

A Deep Learning Approach to the Inversion of Borehole Resistivity Measurements

M. Shahriari · D. Pardo · A. Picon · A. Galdran · J. Del Ser ·
C. Torres-Verdín

Received: date/ Accepted: date

Abstract Borehole resistivity measurements are routinely employed to measure the electrical properties of rocks penetrated by a well and to quantify the hydrocarbon pore volume of a reservoir. Depending on the degree of geometrical complexity, inversion techniques are often used to estimate layer-by-layer electrical properties from measurements. When used for well geosteering purposes, it becomes essential to invert the measurements into layer-by-layer values of electrical resistivity in real time. We explore the possibility of using deep neural networks (DNNs) to perform rapid inversion of borehole resistivity measurements. Accordingly,

we construct a DNN that approximates the following inverse problem: given a set of borehole resistivity measurements, the DNN is designed to deliver a physically reliable and data-consistent piecewise one-dimensional layered model of the surrounding subsurface. Once the DNN is constructed, we can invert borehole measurements in real time. We illustrate the performance of the DNN for inverting logging-while-drilling (LWD) measurements acquired in high-angle wells via synthetic examples. Numerical results are promising, although further work is needed to achieve the accuracy and reliability required by petrophysicists and drillers.

Keywords logging-while-drilling (LWD), resistivity measurements, real-time inversion, deep learning, well geosteering, deep neural networks.

M. Shahriari
Euskampus Fundazioa, Leioa, Spain.
Basque Center for Applied Mathematics, (BCAM), Bilbao, Spain.
Mazarredo 14, E48009 Bilbao, Spain.
E-mail: m.shahriari.sh@gmail.com.com

D. Pardo
University of the Basque Country (UPV/EHU), Leioa, Spain.
Basque Center for Applied Mathematics, (BCAM), Bilbao, Spain.
Ikerbasque (Basque Foundation for Sciences), Bilbao, Spain.
E-mail: dzubiaur@gmail.com

A. Picon
Tecnalia, Bilbao, Spain.
University of the Basque Country (UPV/EHU), Leioa, Spain.
E-mail: artzaip@gmail.com

A. Galdran
École de Technologie Supérieure, Montreal, Canada.
E-mail: adrian.galdran-cabello.1@ens.etsmtl.ca

J. Del Ser
University of the Basque Country (UPV/EHU), Leioa, Spain.
Tecnalia, Bilbao, Spain.
E-mail: javier.delsar@tecnalia.com

C. Torres-Verdín
The University of Texas at Austin, USA. E-mail: cverdin@austin.utexas.edu

1 Introduction

One of the purposes of geophysical measurements is to interrogate the subsurface of the Earth to find oil and gas, and to optimize the production of existing hydrocarbon reservoirs. We divide existing geophysical measurements into two categories: (a) surface geophysical measurements, such as controlled source electromagnetics (CSEM) (see, e.g., [1, 2]), seismic (see, e.g., [3]), and magnetotellurics (MT) (see, e.g., [4]), and (b) borehole sensing, such as logging-while-drilling (LWD) data (see, e.g., [5, 6]).

In this paper, we focus on borehole resistivity measurements. In particular, on those acquired with LWD instruments, which are currently widely used for well geosteering applications (see Figure 1). These logging instruments are equipped with one or various transmitters that emit electromagnetic waves, which are recorded at receivers that are also mounted on the same logging

device. By adequately interpreting (inverting) these measurements, it is possible to determine the subsurface electromagnetic properties nearby the well, thus enabling the selection of an optimal well trajectory to target hydrocarbon-producing zones.

From the mathematical point of view, we identify two different problems depicted in Figure 2:

- **Forward problem:** Given a transmitter t and known material properties (in our case, a resistivity distribution and the geometrical characteristics of the media represented by vector \mathbf{p}), the forward problem delivers the magnetic field (or a post-processed quantity of it) denoted by \mathbf{m} (a vector of measurement) at a receiver r . Denoting by \mathbf{T} a well trajectory composed of several logging positions (i.e., $\mathbf{T} = \{\mathbf{t}_i\}_{i=1}^T$, where T is the number of logging positions), we have:

$$\mathbf{M} = \mathcal{F}(\mathbf{p}; \mathbf{T}), \quad (1)$$

where \mathcal{F} accounts for a partial differential equation (PDE) based on Maxwell's equations and boundary conditions governing the electromagnetic wave propagation phenomena, and $\mathbf{M} = \{\mathbf{m}_i\}_{i=1}^M$ is the vector of measurements acquired along the well trajectory \mathbf{T} , where M is the number of measurements (see e.g., [5, 7, 8]).

- **Inverse problem:** Given a set of measurements \mathbf{M} obtained over a specified logging trajectory \mathbf{T} , the solution of the inverse problem delivers a material subsurface distribution $\mathbf{p} \in \mathbb{R}^P$ (see, e.g., [6, 9, 10]), where P is the number of Parameters characterizing the media. An analytical expression of the governing operator \mathcal{I} that relates these variables is unknown. Nonetheless, for convenience, we express this problem as:

$$\mathbf{p} = \mathcal{I}(\mathbf{M}; \mathbf{T}). \quad (2)$$

Mathematically speaking, the above operator \mathcal{I} is not well-defined. For a given set of input parameters, it may have no output or, as it occurs more frequently, it can provide multiple outputs. These well-known undesirable properties of inverse problems (see, e.g., [11, 12]) make them much more difficult to treat than forward problems. Various techniques such as regularization are intended to overcome these challenges and simplify the solution of inverse problems. The incorporation of non-linear constraints into \mathcal{I} is also a common technique to prevent non-physical solutions (see, e.g., [11]).

An inversion problem is mathematically posed as a minimization of a cost function. There exist multiple

approaches in the literature to solve such minimization problems. A popular one is based on the use of gradient-based algorithms [11, 12]. However, they only guarantee a local minimum, which can be far away from the global one. Another family of methods is based on statistical algorithms [11, 12, 13]. However, they often require a larger number of forward simulations compared to gradient-based methods, which increases the computational time. Moreover, for each new dataset of measurements, one needs to repeat the entire inversion process, which could be computationally intensive. This occurs because none of these methods deliver a full approximation of operator \mathcal{I} itself, but rather they evaluate it over a particular set of measurements.

To decrease the *online* computational time during field operations, we propose to approximate the inverse operator \mathcal{I} *offline* (i.e., *a priori*) using a deep neural network (DNN). Then, the *online* computations reduce to evaluating this approximation for each given set of measurements \mathbf{M} .

In this work, due to the complexity of the inverse problem in hand and the high capacity of DNNs, we focus on using DNNs to approximate our inverse operator. Moreover, by comparing our approach to some common machine learning (ML) algorithms, we prove the superiority of using DNNs for this problem compared to others such as least squares linear regression (LSLR) and random forests (RF).

Early DNNs were already proposed in 1965 [14]. The term *deep learning* was introduced in 1986 [15], and later in 2000 [16] to refer to neural networks (NNs) that contain a large number of layers [17]. A DNN enables to automatically detect and extract complex features that may be present in a given dataset. This was not possible with traditional NNs. In the last decade, DNNs have proven to be useful in multiple areas of knowledge (including computer vision [17], speech recognition [18], and biometrics [19]) to approximate complex functions with unknown properties. In recent years, the use of ML algorithms [20, 21, 22, 23, 24, 25] and deep learning [26, 27] in computational mechanics and computational geophysics has become an active area of study. However, to the best of our knowledge, deep learning algorithms have not been applied to the inversion of borehole resistivity measurements, and therefore, its advantages and limitations on this area are unexplored.

In this work, we provide an introduction for geophysicists on the use of DNNs for solving inverse problems and analyze their main features and limitations when applied to the rapid interpretation of borehole resistivity measurements for geosteering purposes. To simplify the problem and increase the speed of computations, we restrict to Earth formations composed by a

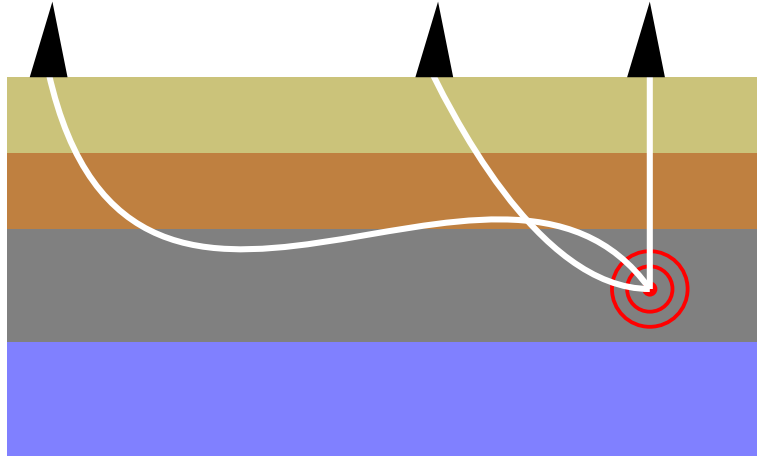


Fig. 1: Drawing of three wells showing how different well trajectories can be employed to reach a specific subsurface target.

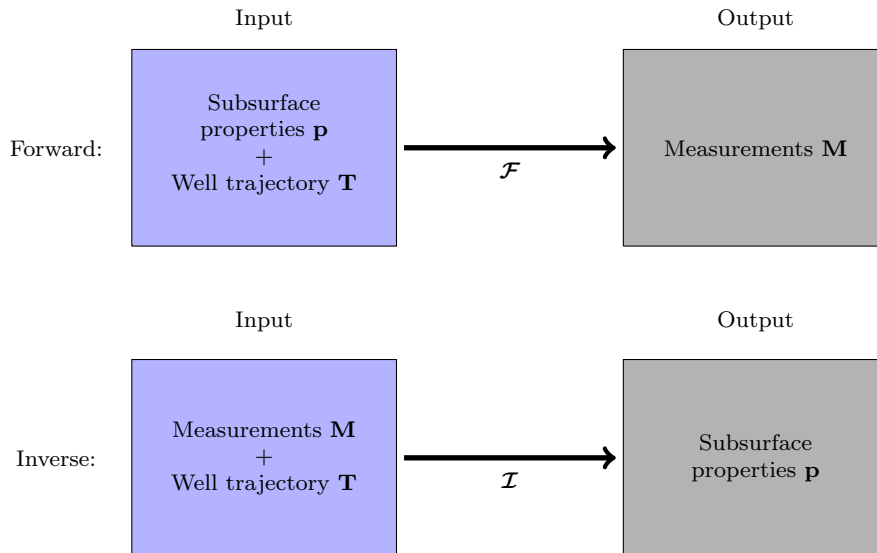


Fig. 2: High-level description of a forward and an inverse problem.

sequence of one-dimensional (1D) layers, as described in [9]. The use of this assumption is common in the oil & gas industry for the inversion of borehole resistivity measurements [6, 28]. However, using only one-dimensional layers, we lose the ability to properly approximate possibly existing geological faults. Hence, we may obtain poor approximations when our instrument is navigating through a geological fault, and also when the LWD instrument is moving from a given 1D section to a different one. These limitations should be taken into account when interpreting the results.

Moreover, for the sake of simplicity, we consider noise-free data samples to train our DNNs. We acknowledge the fact that this is not the case in real-life applications. Nonetheless, we assume this to simplify and

better explore the possibility of using DNNs in the inversion of borehole resistivity measurements as this is the first attempt to do so to the best knowledge of the authors.

To obtain a fair approximation of the inverse operator \mathcal{I} , the dataset which we use to train our DNN should adequately sample the entire data space. How to optimally sample it is a challenging problem that will be the subject of future studies. In here, as a first approximation to tackle this problem, we simply evaluate the performance of our proposed DNN using different sizes of our dataset.

The remaining part of this document is organized as follows. Section 2 provides an introduction to deep learning algorithms. Section 3 describes the governing

equation for borehole resistivity measurements. We introduce our measurement acquisition system in Section 4. Section 5 explains the parameterization (discretization) we select for the well trajectory. A similar description for the material properties discretization is provided in Section 6. Section 7 describes the training of our DNNs, and it shows the results of the training stage. Section 8 demonstrates the applicability of DNNs for inversion of borehole resistivity measurements via synthetic examples. Finally, Section 9 is devoted to conclusions and future work. We also include three appendices that describe some advanced technical details about the DNN employed in this work.

2 Deep Neural Networks for Inverting Resistivity Measurements

In this section, we consider a discrete representation of the inverse function $\mathcal{I}_h : \mathbb{R}^M \times \mathbb{R}^{3T} \rightarrow \mathbb{R}^P$ that associates each pair of measurements and trajectories $(\mathbf{M} \times \mathbf{T}) \in \mathbb{R}^M \times \mathbb{R}^{3T}$ with a corresponding distribution of subsurface properties $\mathbf{p} \in \mathbb{R}^P$. In order to approximate this function, we employ NNs [29]. We provide below a concise overview of how to construct this kind of operators. The existing literature about NNs is large, but in here we only intend to briefly introduce some NNs and related algorithms to geophysicists that are relevant for the inversion of borehole resistivity measurements.

2.1 Fully-Connected Neural Network

Early formulations of NNs, known as fully-connected neural networks (FC-NNs), were defined by repeated compositions of simple transformations. Denoting $\mathbf{x} = (\mathbf{M}, \mathbf{T})$, an FC-NN composed of L layers is given by:

$$\mathcal{I}_\theta(\mathbf{x}) = (\mathcal{N}^{(L)} \circ \dots \circ \mathcal{N}^{(l)} \circ \dots \mathcal{N}^{(2)} \circ \mathcal{N}^{(1)})(\mathbf{x}), \quad (3)$$

where $\mathcal{N}^{(l)}(\mathbf{x}) = \mathbf{s}(\mathbf{W}^{(l)} \cdot \mathbf{x} + \mathbf{b}^{(l)})$, $\mathbf{W}^{(l)}$ is a matrix, and $\mathbf{b}^{(l)}$ a vector. Thus, $\mathbf{W}^{(l)} \cdot \mathbf{x} + \mathbf{b}^{(l)}$ is an affine transformation. \mathbf{s} is a simple non-linear point-wise mapping (activation function), typically the so-called *rectified linear unit* given by:

$$\mathbf{s}(v_1, \dots, v_r) = (\max(0, v_1), \dots, \max(0, v_r)). \quad (4)$$

Other activation functions can be applied, with arguably worse gradient preserving properties (e.g., tanh). We define $\boldsymbol{\theta}^{(l)}$ as a vector composed of all entries of matrix $\mathbf{W}^{(l)}$ and vector $\mathbf{b}^{(l)}$ for each layer $l = 1, \dots, L$. Thus, $\boldsymbol{\theta} = \{\boldsymbol{\theta}^{(l)} : 1 \leq l \leq L\}$ is a large vector of parameters fully determining \mathcal{I}_θ . Due to the varying dimensions

of the different matrices $\mathbf{W}^{(l)}$ and vectors $\mathbf{b}^{(l)}$ at each layer in Equation (3), the dimensionality of the input \mathbf{x} can change, eventually reaching that of the target variable $\mathbf{p} \in \mathbb{R}^P$.

2.2 Training an NN: Data Preparation

We consider a finite set \mathcal{S} containing m data samples:

$$\mathcal{S} = (\mathcal{M}, \mathcal{T}, \mathcal{P}) = \{(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}, \mathbf{p}^{[i]}) : \mathbf{M}^{[i]} \in \mathbb{R}^M, \mathbf{T}^{[i]} \in \mathbb{R}^{3T}, \mathbf{p}^{[i]} \in \mathbb{R}^P, 1 \leq i \leq m\}. \quad (5)$$

This set is randomly split into three disjoint subsets, referred to as *training*, *validation*, and *test* sets, respectively:

$$\begin{aligned} \mathcal{S}_{train} &= (\mathcal{M}_{train}, \mathcal{T}_{train}, \mathcal{P}_{train}) \\ &= \{(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}, \mathbf{p}^{[i]}) : 1 \leq i \leq m_1\}, \\ \mathcal{S}_{val} &= (\mathcal{M}_{val}, \mathcal{T}_{val}, \mathcal{P}_{val}) \\ &= \{(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}, \mathbf{p}^{[i]}) : m_1 + 1 \leq i \leq m_2\}, \\ \mathcal{S}_{test} &= (\mathcal{M}_{test}, \mathcal{T}_{test}, \mathcal{P}_{test}) \\ &= \{(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}, \mathbf{p}^{[i]}) : m_2 + 1 \leq i \leq m\}. \end{aligned} \quad (6)$$

We apply a network \mathcal{I}_θ to the input data sampled from set \mathcal{S} in order to produce a prediction $\mathcal{I}_\theta(\mathbf{M}^{[i]}, \mathbf{T}^{[i]})$ of its resistivity values. Then, one can compute the accuracy of such prediction via an error function \mathcal{L} , in our case given by the l_1 norm of the difference between both vectors:

$$\mathcal{L}(\mathcal{I}_\theta(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}), \mathbf{p}^{[i]}) = \|\mathcal{I}_\theta(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}) - \mathbf{p}^{[i]}\|_1. \quad (7)$$

The numerical process by which the error given by Equation (7) is iteratively minimized via a gradient-based algorithm across the entire training set is referred to as *training*, and will be detailed in the next subsection. The validation set \mathcal{S}_{val} is employed to perform some high-level NN design decisions, e.g., to modify the network architecture (the dimensions of the different layers) or different parameters controlling the numerical optimization algorithm. After training, we compute predictions for data samples in \mathcal{S}_{test} . Then, the network \mathcal{I}_θ is said to generalize properly if the errors in \mathcal{S}_{train} and \mathcal{S}_{test} are similar. In addition, if such errors are relatively low, we can assume that \mathcal{I}_θ correctly approximates operator \mathcal{I}_h .

2.3 Training an NN: Numerical Optimization

A critical feature of NNs is that they are a hierarchical composition of multiple functions that are easy to

differentiate. Hence, the chain rule becomes an essential tool to find derivatives of these operators. This is the core idea of the most popular algorithm for implementing gradient descent strategies on NNs, called *back-propagation* in the NN's literature [30].

Within each gradient descent iteration, we first carry out a forward pass for a given data sample (m_i, t_i) in order to compute a prediction $\mathcal{I}_\theta(\mathbf{M}^{[i]}, \mathbf{T}^{[i]})$ and the corresponding error $\mathcal{L}(\mathcal{I}_\theta(\mathbf{M}^{[i]}, \mathbf{T}^{[i]}), \mathbf{p}^{[i]})$. Afterwards, this error is backpropagated by applying the chain rule to the composition of functions defining the different layers of the network. Hence, proceeding from the last layer of the model backwards, one can estimate the gradient of the loss function with respect to parameters θ , defining \mathcal{I}_θ in reverse order as:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta^{(L)}}, \dots, \frac{\partial \mathcal{L}(\theta)}{\partial \theta^{(l+1)}}, \frac{\partial \mathcal{L}(\theta)}{\partial \theta^{(l)}}, \dots, \frac{\partial \mathcal{L}(\theta)}{\partial \theta^{(1)}} \quad (8)$$

The gradient at each layer is derived based on previous gradient computations, parameters θ are updated with some form of gradient descent strategy (e.g., stochastic gradient descent), and the process is iterated over all elements of the training set so as to minimize its average error.

The number of iterations during which the model is trained is typically decided by monitoring the value of the loss function \mathcal{L} on elements of the validation set \mathcal{S}_{val} that are never used to adjust the network parameters. During training, that value is compared with the loss value attained in \mathcal{S}_{train} in order to stop the optimization process as soon as both quantities start to diverge, which would imply that the network is becoming too much adjusted to the training data and failing to generalize, a phenomenon known as *overfitting*.

2.4 Convolutional Neural Networks

As observed from Equation (3), NNs are defined as a composition of functions. Thus, they naturally possess a layer-wise hierarchical nature. Therefore, they are ideal candidates to design operators that progressively retain the most salient aspects of the initial input. However, $\mathbf{W}^{(l)}$ are dense matrices, connecting every component of the input of a given layer to its output. This results in an excessively large number of parameters that need to be optimized. In order to reduce this number, a popular solution consists of replacing fully-connected affine layers \mathcal{N} by convolutional operators \mathcal{C} defined by convolution kernels f . This localizes computations, effectively reducing the number of parameters in \mathcal{I}_θ . The resulting network is known as a convolutional neural network (CNN) [31]. We provide a rigorous definition of a CNN in Appendix A.

2.5 Recurrent Neural Networks

A particular kind of network architectures that are useful for sequence processing (e.g., speech, text, or time-related data) are recurrent neural networks (RNNs) [32]. In here, since successively recorded logging data exhibits a temporal pattern (there is a strong relationship between measurements recorded at a given logging position and at subsequent ones), we will also adopt an RNN design. For a technical description of this type of networks, see Appendix B.

2.6 An NN Architecture for Inverting Borehole Resistivity Measurements

Selecting the optimal NN architecture for any problem is still an open research area. However, an analysis of the basic features of the data and capability of each type of NN provides some intuitions that can serve as a guide to construct a suitable DNN architecture.

In this work, since the measurements are recorded for a sequence of logging positions, we can interpret them in the form of a time series. Therefore, we first consider a specific class of RNN named long short-term memory (LSTM) to reduce the dimensionality of the data and to extract the features between the measurements for each logging position and the measurements of the previous and future logging positions (see Appendix B). Then, the output of the above LSTM layer serves as the input of a series of one-dimensional CNNs, with interleaved *pooling* operators similar to the ones described in Appendix A, where each convolutional block is based on a modified residual block [33] (see Appendix A). The above mentioned CNNs layer helps us to extract a useful data representation of the data while reducing their dimensionality. The interleaved *pooling* operators further help in reducing data dimensionality while preserving the most salient data features. In the end, the output of this second set of operations becomes the input to a fully-connected layer that maps it into space \mathbb{R}^P of subsurface properties. This fully-connected layer performs the final and most effective feature extraction on the dimensionality-reduced data provided by the previous NN layers. The network is trained end-to-end by backpropagation until the validation error is no longer decreasing. We provide a pseudo-code of this DNN in Appendix C.

By considering the above-explained NN, we achieved the results with high R^2 factor presented in Section 7.3. We remark that the architecture of the NN is not unique. The selected architecture is the result of a series of design choices that are made after multiple experiments with different configurations. In each of these

experiments, a different model is optimized based on the available training data, and the performance on the validation set is monitored to decide about the most suitable architecture design for our problem. Investigating different NN architectures is the subject of future studies.

3 Reduced Wave Magnetic Field Equation

Let \mathbf{H} be the magnetic field, \mathcal{M} a magnetic source flux density, and $\boldsymbol{\sigma} = \boldsymbol{\rho}^{-1}$ a real-valued conductivity tensor with positive determinant. Then, the following reduced wave equation governs the magnetic field propagation phenomena:

$$\nabla \times \tilde{\boldsymbol{\sigma}}^{-1} \nabla \times \mathbf{H} - i\omega \boldsymbol{\mu} \mathbf{H} = i\omega \boldsymbol{\mu} \mathcal{M}, \quad (9)$$

where $\tilde{\boldsymbol{\sigma}}^{-1} = (\boldsymbol{\sigma} - i\omega \boldsymbol{\varepsilon})^{-1}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\mu}$ are the permittivity and magnetic permeability tensors of the media, respectively, $\omega = 2\pi f$ is the angular frequency, where $f > 0$ is the frequency of operation of the transmitter, and i is the imaginary unit, $i^2 = -1$. The problem domain is $\Omega = \mathbb{R}^3$.

In this work, we consider a sequence of 1D transversally isotropic (TI) media [9]. Therefore, the resistivity of the media varies only along z -direction (see Figure 3), and we have:

$$\boldsymbol{\rho}(z) = \begin{pmatrix} \rho_h(z) & 0 & 0 \\ 0 & \rho_h(z) & 0 \\ 0 & 0 & \rho_v(z) \end{pmatrix}, \quad (10)$$

where ρ_h and ρ_v are the horizontal and vertical resistivities of the media, respectively.

4 Measurement Acquisition System

In this work, we first consider the short co-axial LWD instrument as shown in Figure 4. For this instrument, we measure attenuation and phase difference. We will denote those measurements as \mathcal{M}_1 . To compute them, we consider the zz coupling H_{zz} , where the first and the second subscripts correspond to the direction of the transmitter and the receiver, respectively. We record these quantities at both receivers, and denote them as H_{zz}^1 and H_{zz}^2 . We define the attenuation and the phase difference as follows:

$$\ln \frac{H_{zz}^1}{H_{zz}^2} = \underbrace{\ln \frac{|H_{zz}^1|}{|H_{zz}^2|}}_{\times 20 \log(e) = \text{attenuation (dB)}} + i \underbrace{(ph(H_{zz}^1) - ph(H_{zz}^2))}_{\times \frac{180}{\pi} = \text{phase difference (degree)}}, \quad (11)$$

where ph denotes the phase of a complex number.

In addition, we consider the short-spacing deep azimuthal instrument described in Figure 5. For this logging instrument, we record the attenuation and phase difference and denote these measurements as \mathcal{M}_2 . We define them as in Equation (11) with $H_{zz}^2 = 1$ since there is no second transmitter. Finally, we also record a directional measurement referred as geosignal and defined as follows:

$$g = \ln \frac{H_{zz} - H_{zx}}{H_{zz} + H_{zx}} = \underbrace{\ln \frac{|H_{zz} - H_{zx}|}{|H_{zz} + H_{zx}|}}_{\times 20 \log(e) = \text{attenuation (dB)}} + i \underbrace{(ph(H_{zz} - H_{zx}) - ph(H_{zz} + H_{zx}))}_{\times \frac{180}{\pi} = \text{phase difference (degree)}}. \quad (12)$$

\mathcal{M}_3 denotes the set of geosignal measurements.

5 Trajectory Parameterization

We select a fixed number of tool positions based on the depth of investigation of the logging instruments. For our instruments, the largest depth of investigation is close to 20 m. By considering the logging step size equal to one foot (0.3048 m), we select $T = 65$.

We consider an arbitrary (but close to horizontal) trajectory, as it is customary in geosteering applications. Since we assume a 1D layered media on the proximity of the well trajectory, we select the azimuthal degree of the trajectory to be always equal to zero.

With the above assumptions, we discretize (parameterize) the well trajectory as follows. We consider $\alpha_{ini}(\mathbf{t})$ to be the initial trajectory dip angle. We assume that the trajectory dip angle can vary while drilling by an angle α_v in each step. Hence, at each tool position (i), the trajectory dip angle is:

$$\alpha(\mathbf{t}_i) = \alpha_{ini}(\mathbf{t}) + (i - 1)\alpha_v, \quad i = 1, \dots, T, \quad (13)$$

where $\alpha(\mathbf{t}_i)$ is the trajectory dip angle at the i -th position.

6 Material Properties Parameterization

For the successful 1D inversion of borehole resistivity measurements, it is often sufficient to recover a media containing only three layers for each logging position and it is characterized by the following seven variables: (A) the horizontal and vertical resistivity of the layer where the tool is currently located (ρ_h and ρ_v , respectively); (B) the resistivity of the assumed isotropic upper and lower layers located above and below the current logging position (ρ_u and ρ_l , respectively); (C) the

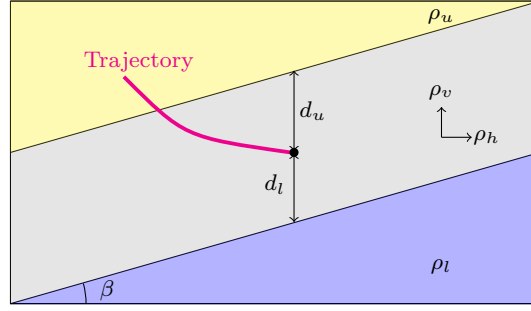


Fig. 3: 1D media and a trajectory. The black circle indicates the last position of the trajectory.

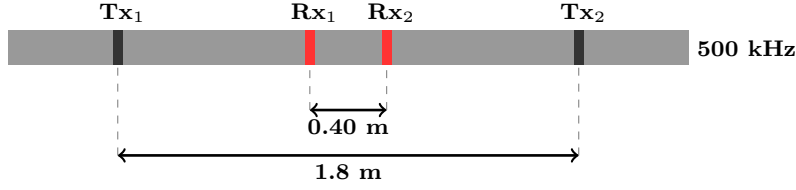


Fig. 4: Conventional LWD logging instrument. Tx_i and Rx_i are the transmitters and the receivers, respectively.

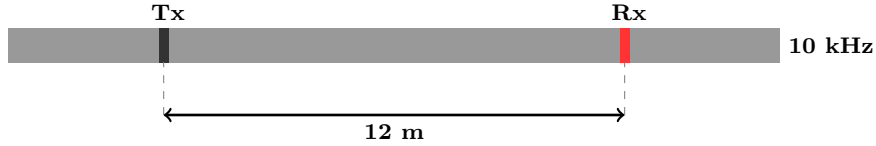


Fig. 5: Short-spacing deep azimuthal instrument. Tx and Rx are the transmitter and the receiver, respectively.

vertical distance from the current logging position to the upper and lower bed boundary positions (d_u and d_l , respectively); and (D) the dip angle of the formation (β), which is assumed to be identical for all layers (see Figure 3). Our DNN will provide an estimate of these seven numbers at each logging position.

7 Training the DNN

To produce reliable training and validation sets, and to avoid full randomness which may lead to non-physical data, we consider some physical and geological properties of the subsurface.

7.1 Material properties

In order to produce our training and validation sets, we select $\rho_u, \rho_l \in [1, 10^3]$, which is a physical restriction observed in most geological structures. Since we want resistivity values to be comparable, we consider them in logarithmic scale. Thus, our random variables become: $\log(\rho_l), \log(\rho_u) \in [0, 3]$. Additionally, in the case of ρ_h

and ρ_v , to impose a physically meaningful structure of the subsurface, we incorporate the following physical restrictions:

$$\rho_h \leq \rho_v \leq 10\rho_h. \quad (14)$$

Therefore, we obtain:

$$1 \leq \underbrace{\frac{\rho_v}{\rho_h}}_a \leq 10, \quad (15)$$

where a is the anisotropy factor. In order to impose restriction (14) in our calculations, we select random values of $\log(a) \in [0, 1]$. Moreover, since we want to have $\log(\rho_v) \in [0, 3]$, we select $\log(\rho_h) \in [0, 3 - \log(a)]$. We consider $d_l, d_u \in [0.01, 10]$ meters. As with the resistivity values, we consider them in logarithmic scale, i.e., our variables become $\log(d_l), \log(d_u) \in [-2, 1]$. As it often happens in geological layers, we assume that the dip angle is $\beta \in [-10^\circ, 10^\circ]$.

To summarize, we select $\log(\rho_u), \log(\rho_l), \log(\rho_h), a, \log(d_u), \log(d_l)$, and β randomly within their aforementioned ranges of variation to characterize our synthetic forward models.

7.2 Trajectory

In this work, we consider an almost horizontal trajectory, as it occurs in most geosteering applications. Specifically, we restrict to $\alpha_{ini}(\mathbf{t}) \in [83^\circ, 97^\circ]$. Moreover, we further assume that the tool trajectory deviates by a maximum of 3° in a 20 meters section. In addition, since the direction of the trajectory dip angle is often changing gradually and almost constantly from one logging position to the next, for $T = 65$ we have $\alpha_v \in [-0.045^\circ, 0.045^\circ]$.

By selecting randomly $\alpha_{ini}(\mathbf{t})$ and α_v in their above ranges of variation, we build the trajectories for our forward problems.

7.3 Results

To produce a dataset enabling optimal sampling of the underlying data distribution, we first evaluated our DNN with different amounts of data samples, resulting in training datasets of different sizes. Figure 6 compares the R^2 score of our DNN evaluated on a test dataset while using datasets of varying sizes. Based on those results, we generate one million randomly selected samples/trajectories and their associated formation models (80% training, 10% validation, and 10% test). Figure 7 shows the accuracy of the trained DNN when we only consider the set of measurements \mathcal{M}_2 , i.e., $\mathcal{M} = \mathcal{M}_2$. The red line indicates the perfect approximation where the predicted value and the ground truth (the real parameters associated with a formation) coincide. The upper and lower blue lines show percentiles 10 and 90, respectively. These percentiles provide a reliable uncertainty quantification. In a perfect approximation, the blue lines should coincide with the red one. Therefore, a lower distance between the blue lines and the red one indicates a better approximation. Figures show a denser cloud of points in the proximity of the red line, which indicates an acceptable approximation. However, for the anisotropy factor a , the DNN is almost unable to predict the correct value, and consequently, it can not predict ρ_v as precisely as the other variables.

Analogously, Figure 8 illustrates the results when we select $\mathcal{M} = \mathcal{M}_2 \cup \mathcal{M}_3$. One can see that the blue lines are closer compared to those shown in Figure 7. Moreover, the concentration of points in the proximity of the red line increases. However, the approximation of the anisotropy factor a is still poor, although better than in the previous case.

Figure 9 illustrates the results when we employ all available measurements, i.e., $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$. These results outperform previous ones obtained with

fewer measurements and, for the first time, we obtain an acceptable prediction of anisotropy factor a .

To compare the results obtained with a DNN vs those delivered by other ML methods, we consider two additional ML algorithms. In both cases, we use incremental principal component analysis (IPCA) as a preprocessor to decrease the dimensionality of the input data. Then, we use the following two traditional algorithms to approximate the inverse operator: (a) a least squares linear regression (LSLR), and (b) a random forests (RF) of 150 compounding trees and a maximum tree depth of 20. Figure 10 shows the superiority of using DNN for solving our inverse problem compared to the other considered ML approaches by displaying their corresponding R^2 scores.

8 Inversion Results

Since the DNN trained with measurements drawn from $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$ exhibits the best performance, we use it to invert several practical synthetic examples.

Figure 11 illustrates the inversion of a three-layer media in which the middle layer is more conductive than the other ones, and it is anisotropic. Inversion results are less accurate than those possibly obtained with a gradient-based method. However, as initial results, they are encouraging. The results show that for the isotropic layer, the prediction of the resistivity is better than the one for the anisotropic layer. This probably occurs because the inversion of anisotropic factor a presents some accuracy deficiencies. The predictions of d_u and d_l provide an acceptable view of the material surrounding the instrument. Figures 12 and 13 show a comparison between the attenuations and the phase differences of the measurements corresponding to the exact and predicted (inverted) models. These results show a better approximation of \mathcal{M}_1 than of \mathcal{M}_2 and \mathcal{M}_3 . This noise may be related to the inverse problem not being well defined, i.e. the existence of multiple outputs for a given input. One standard option to minimize the noise in inverse problems is to incorporate some regularization terms to further approximate our solution to a specific subset of possible outputs. We add the following standard regularization term for the inversion of resistivity measurements to the cost function:

$$\mathcal{R}_T = \alpha \|\nabla \mathbf{p}\|_1, \quad (16)$$

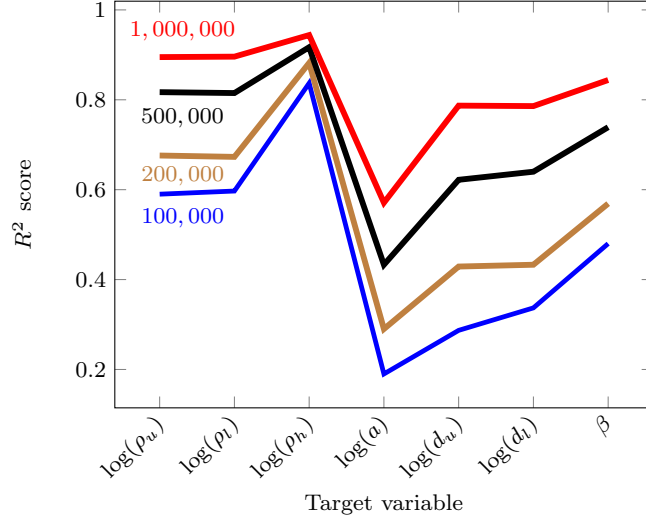


Fig. 6: R^2 scores measured over the test set for our DNN approach for different sizes of dataset samples.

where for $\mathbf{p} = (\log(\rho_u), \log(\rho_l), \log(\rho_h), \log(a), \log(d_u), \log(d_l), \beta)$ we define:

$$\begin{aligned}
 (\nabla \mathbf{p})_1 &= \log(\rho_u) - \log(\rho_h), \\
 (\nabla \mathbf{p})_2 &= \log(\rho_l) - \log(\rho_h), \\
 (\nabla \mathbf{p})_3 &= \log(\rho_h) - 0.5 \log(\rho_u) - 0.5 \log(\rho_l), \\
 (\nabla \mathbf{p})_4 &= \log(a), \\
 (\nabla \mathbf{p})_5 &= 0, \\
 (\nabla \mathbf{p})_6 &= 0, \\
 (\nabla \mathbf{p})_7 &= 0.1\beta.
 \end{aligned} \tag{17}$$

The regularization imposed using (17) combined with a sufficiently large regularization parameter α directs us to the one solution of the inverse problem which is closest to a homogeneous media. Figure 14 shows the inversion results when using regularization. We observe that even after adding this regularization term, noise artefacts still take place in the solution. To further minimize noise, we believe it is necessary to design a modified cost function and consider some additional regularization terms. This will be the subject of future studies.

Figure 15 displays an inversion performed on a three-layer media in which the middle layer is isotropic and also the most resistive one. We consider the other two layers to be anisotropic. As in the previous model problem, results show discrepancies in the anisotropic layers probably because of the lack of a good approximation of anisotropy factor a . Figures 16 and 17 compare the measurements corresponding to the exact and predicted (inverted) models.

Figure 18 describes the inversion results performed on a synthetic example containing a sequence of 1D lay-

ered media. Each 1D model consists of four layers. Inverted results show visible imperfections, and the lack of accuracy for anisotropy factor a causes a poor approximation of the resistivity value in the anisotropic layer. Predicted d_u and d_l could be employed as a first approximation of the formation surrounding the logging instrument, although a better estimation of d_u and d_l is necessary for a more accurate indicator of the bed boundary positions. Figures 19 and 20 compare the measurements corresponding to the exact and predicted (inverted) models. As in previous results, the best approximation is exhibited in the logs corresponding to \mathcal{M}_1 .

Figures 21, 22, and 23 present inversion results for a new synthetic example. A second trajectory is considered to obtain Figures 24, 25, and 26. Again, even if the results present noticeable inaccuracies, DNN results can be used as fast initial approximations that could be refined with other more expensive methods. Notice that DNN inversion results are obtained in a few seconds for over a thousand logging positions and they also provide an uncertainty map.

9 Discussion and Conclusion

In this work, we investigated the use of deep learning for the inversion of borehole resistivity measurements. The training stage of a DNN can be a time-consuming stage which can take up to three weeks using GPU to obtain a good approximation. However, we perform the training stage *offline*. Then, the *online* stage (actual inversion) of the method is faster than all other existing conventional inversion methods, which makes it ideal

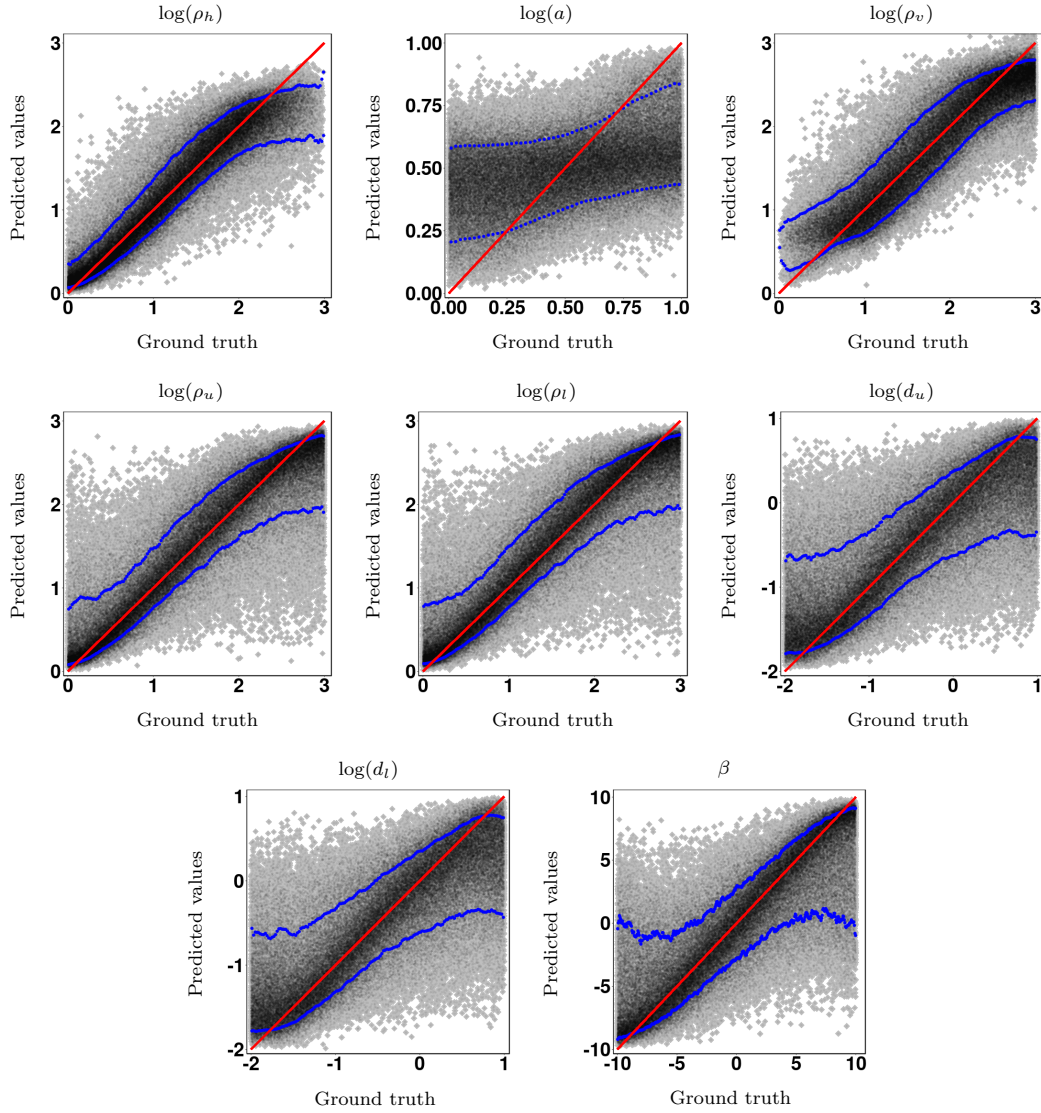


Fig. 7: Model problem 2. Comparison between the ground truth and predicted values using a trained DNN for $\mathcal{M} = \mathcal{M}_2$. Red line indicates the equality of the predicted values and the ground truth. The blue lines correspond to the 10 and 90 percentiles, respectively.

for geosteering purposes. Additionally, using DNNs, we can provide a reliable uncertainty quantification map. Thus, there is excellent potential in using DNNs for this application.

However, DNNs also present important limitations. First, the inverted results shown in this exploratory work present inaccuracies and further research in the area is still needed. Second, in order to train the system, we require a massive number of data. In the case of a 1.5D model problem, rapid forward solvers exist, which can produce the required data in a reasonable amount of time. However, in the case of 2D and 3D problems, producing such training data set may be extremely time-consuming. Moreover, because of the com-

plexity of the problem and the number of variables in the case of 2D and 3D model problems, a much larger data set is required compared to the case of 1.5D problems. Hence, further research is necessary in order to successfully apply DNNs for the inversion of 2D and 3D problems. Third, exploring all possible venues and producing a reliable inversion method using DNNs requires a considerable amount of computational resources and prospective design experimentation. Fourth, the understanding of deep learning algorithms is still limited. In particular, it does not exist a mathematically sound algorithm for the optimal design of the best possible DNN for a given problem. Similarly, it is difficult to recognize a poorly designed DNN. Finally, another limitation of

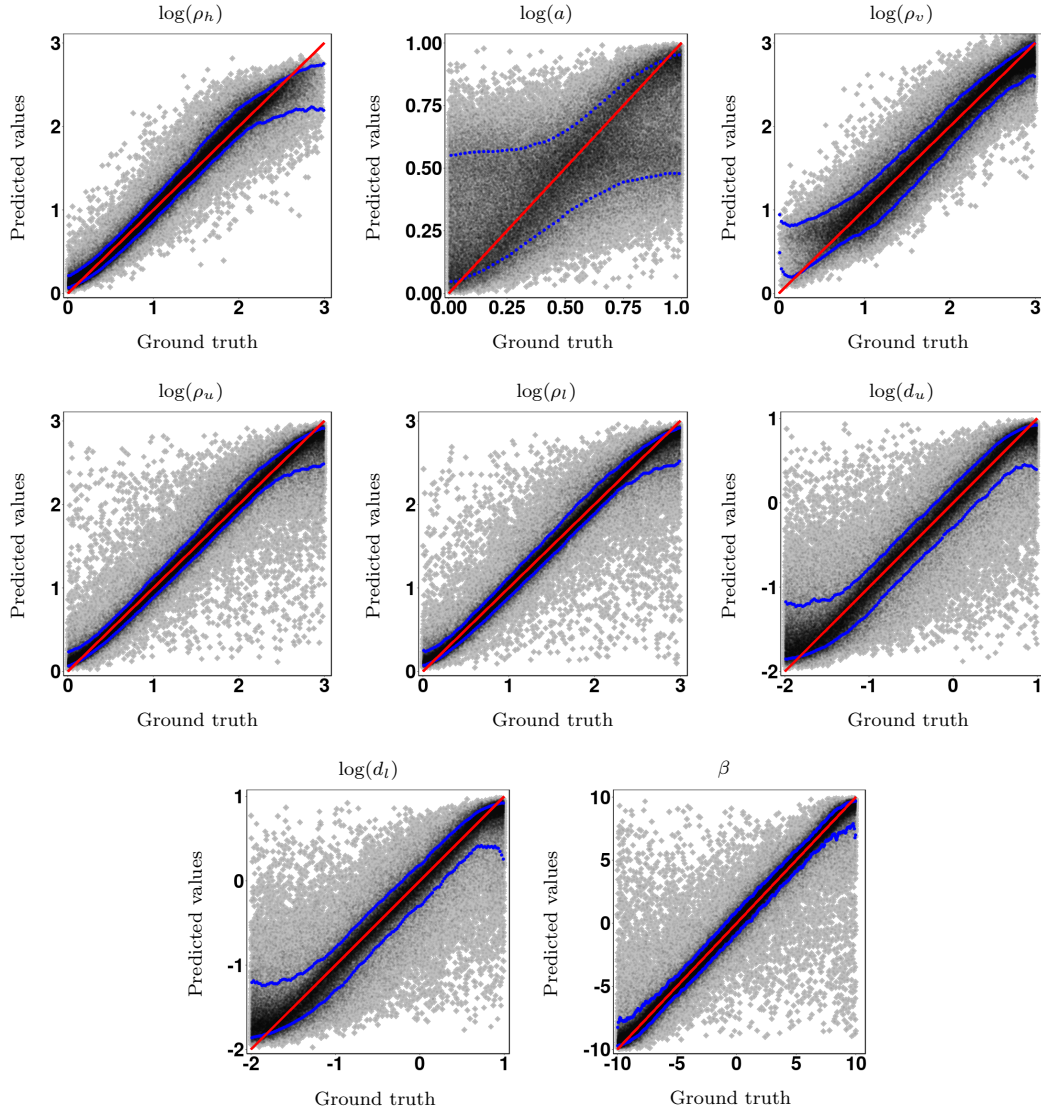


Fig. 8: Model problem 2. Comparison between the ground truth and predicted values using a trained DNN using $\mathcal{M} = \mathcal{M}_2 \cup \mathcal{M}_3$. Red line indicates the equality of the predicted values and the ground truth. The blue lines correspond to the 10 and 90 percentiles, respectively.

DNNs is that they can only compute a discrete version of the inversion function and when modifying the dimensionality of the measurements, a new DNN should be designed.

The results presented in this work are promising. However, extensive work is still needed in the field to achieve the required accuracy. We envision a large area of research on the topic. As future work, we want to produce more advanced DNNs by designing mechanisms to embed physical constraints associated with our problem into their construction. Furthermore, we want to investigate the use of DNNs for the design of measurement acquisition systems. We can use DNNs for each instrument configuration we design and observe the sen-

sitivity of the desired design to the inversion variables. As a natural step towards solving industrial application problems, we shall also investigate the accuracy of the DNNs for noisy data and we will include regularization terms in the cost functional. Moreover, we shall work on the design of employing different cost functions and regularization terms to minimize the error. The use of other ML algorithms also seems an interesting path to follow.

Acknowledgements Mostafa Shahriari and David Pardo have received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 777778, the Projects

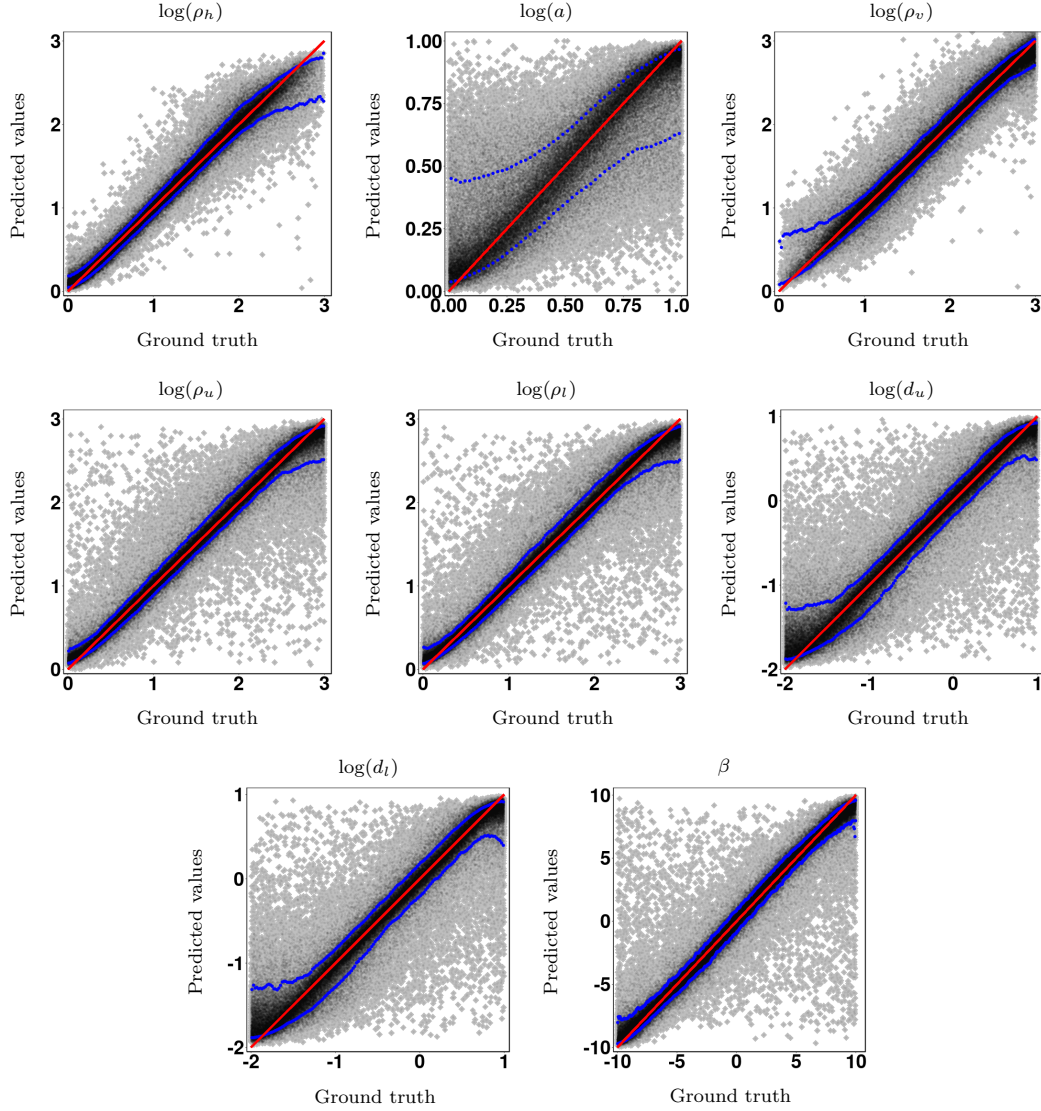


Fig. 9: Model problem 2. Comparison between the ground truth and predicted values using a trained DNN using $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$. Red line indicates the equality of the predicted values and the ground truth. The blue lines correspond to the 10 and 90 percentiles, respectively.

of the Spanish Ministry of Economy and Competitiveness with reference MTM2016-76329-R (AEI/FEDER, EU), and MTM2016-81697-ERC/AEI, the BCAM “Severo Ochoa” accreditations of excellence SEV-2013-0323 and SEV-2018-0718, and the Basque Government through the BERC 2014-2017 program, and the Consolidated Research Group Grant IT649-13 on “Mathematical Modeling, Simulation, and Industrial Applications (M2SI)”.

Javier Del Ser thanks the Basque Government for its funding support through the EMAITEK program.

Carlos Torres-Verdín was partially funded by The University of Texas at Austin Research Consortium on Formation Evaluation, jointly sponsored by AkerBP, Anadarko, Aramco, Baker-Hughes, BHP Billiton, BP, China, Oilfield Services LTD., Chevron, ConocoPhillips, DEA, ENI, Equinor ASA, Halliburton, Inpex, Lundin-Norway, Nexen, Petrobras, Repsol, Shell,

Southwestern, TOTAL, Wintershall and Woodside Petroleum Limited.

Appendices

A Convolutional Neural Networks

convolutional neural networks (CNNs) [31] are a particular kind of NNs built by replacing fully-connected affine layers \mathcal{N} by convolutional operators \mathcal{C} defined by convolution kernels \mathbf{f} . Hence, Equation (3) becomes:

$$\mathcal{I}_\theta(x) = (\mathcal{C}^{\mathbf{f}^{(L)}} \circ \dots \circ \mathcal{C}^{\mathbf{f}^{(l)}} \circ \dots \circ \mathcal{C}^{\mathbf{f}^{(2)}} \circ \mathcal{C}^{\mathbf{f}^{(1)}})(x), \quad (18)$$

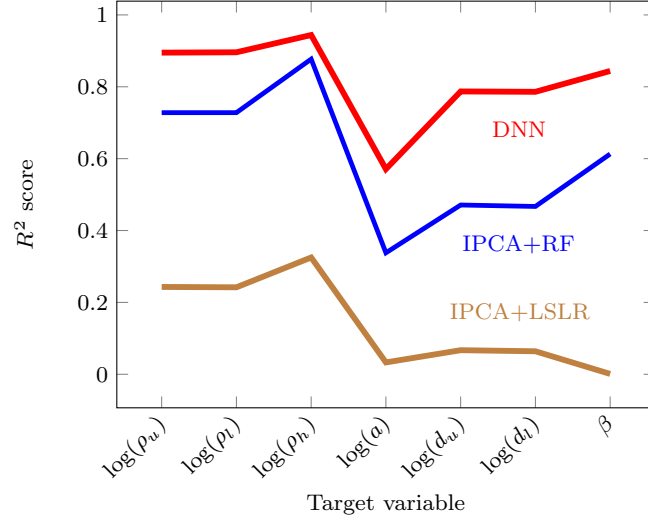


Fig. 10: R^2 scores measured over the test set for our DNN approach and two ML approaches: (a) combining IPCA and RF, and (b) combining IPCA and LSLR.

In a discrete setting, at layer l of Equation (18), operator $\mathcal{C}^{(l)}$ is determined by the set of convolutional kernels $\mathbf{f}^{(l)} = \{\mathbf{f}_s^{(l)}, s = 1, \dots, c_{l+1}\}$. Each of these kernels transforms an input tensor $\mathbf{x}^{(l)}$ of dimension $h_l \times w_l \times c_l$ into an output $\mathbf{x}_s^{(l+1)}$ of dimension $h_l \times w_l$. Each kernel is defined by a tensor of dimension $M_l \times N_l \times c_l$ that acts on its inputs through a simple convolution-like operation, followed by a non-linear function like the one in Equation (4):

$$\mathbf{x}_s^{(l+1)}(h, w) = \mathbf{s} \left(\sum_{m=1}^{M_l} \sum_{n=1}^{N_l} \sum_{c=1}^{c_l} \mathbf{f}_s^{(l)}(m, n) \cdot \mathbf{x}^{(l)}(h+m, w+n, c) \right). \quad (19)$$

Application of all the c_{l+1} convolution kernels of $\mathbf{f}^{(l)}$ on the input $\mathbf{x}^{(l)}$ finally results into an output tensor $\mathbf{x}^{(l+1)}$ of dimension $h_l \times w_l \times c_{l+1}$. Each of these convolutional layers $\mathcal{C}^{(l)}$ is followed by a non-linear point-wise function, and the spatial size of the output from each layer is decreased by a fixed projection operator $\mathcal{P}^{(l)} : \mathbb{R}^{h_l \times w_l} \rightarrow \mathbb{R}^{h_{l+1} \times w_{l+1}}$. Typically, $\mathcal{P}^{(l)}$ is defined as a local averaging operation. Again, eventually the dimensionality of the initial input \mathbf{x} is transformed into that of an element of the target space \mathbb{R}^P .

B Recurrent Neural Networks

Let us first consider a simple neural network with an input, an intermediate, and an output layer like the one defined in Section 2.1 as a directed graph in which nodes store the result of the operations described in

Equation (3) and edges store the weights of the network \mathbf{W} , \mathbf{b} , as in Figure 27a. Computations performed by such a network to obtain an output, given an input \mathbf{x} , are described as:

$$\begin{aligned} \mathbf{z}^{(1)} &= \mathbf{s}(\mathbf{a}^{(1)}) = \mathbf{s}(\mathbf{W}^{(1)} \cdot \mathbf{x} + \mathbf{b}^{(1)}), \\ \mathcal{I}_\theta(\mathbf{x}) &= \mathbf{s}(\mathbf{W}^{(2)} \cdot \mathbf{z}^{(1)} + \mathbf{b}^{(2)}), \end{aligned} \quad (20)$$

where $\mathbf{a}^{(1)}$, also known as *activation*, denotes the output of the network at the first layer of this network before passing through the non-linearity \mathbf{s} . The key difference between regular NN and a recurrent neural network (RNN), as shown in Figure 27b, is that the graph defining an NN is acyclical, whereas in an RNN internal cycles are allowed. This introduces a notion of time or sequential dependency into the computations of the network.

In our case, we interpret a data sample as a temporal sequence of length T , $\mathbf{x} = (x_1, x_2, \dots, x_T)$, and the goal is to predict an output sequence \mathbf{p} from \mathbf{x} . In an RNN, a regular NN is trained to predict $\mathbf{p} = \mathcal{I}_\theta(\mathbf{x}_t)$ out of \mathbf{x}_t for $1 \leq t \leq T$, but the data is scanned left-to-right, and the previous activation is multiplied by a second set of learnable weights. Hence, the necessary computations within an RNN for a forward pass are specified by the following two equations:

$$\begin{aligned} \mathbf{a}_t &= \mathbf{W}_{\mathbf{ax}} \mathbf{x}_t + \mathbf{W}_{\mathbf{aa}} \mathbf{a}_{t-1} + \mathbf{b}_{\mathbf{a}} \\ \mathcal{I}_\theta(\mathbf{x}_t) &= \mathbf{s}(\mathbf{W}_{\mathbf{pa}} \mathbf{a}_t + \mathbf{b}_{\mathbf{p}}), \end{aligned} \quad (21)$$

where $\mathbf{W}_{\mathbf{ax}}$ is a matrix of conventional weights between the input and the inner layer, $\mathbf{W}_{\mathbf{aa}}$ is a matrix holding recurrent weights between the inner layer at time step t and itself at adjacent time step $t+1$, $\mathbf{W}_{\mathbf{ax}}$ maps

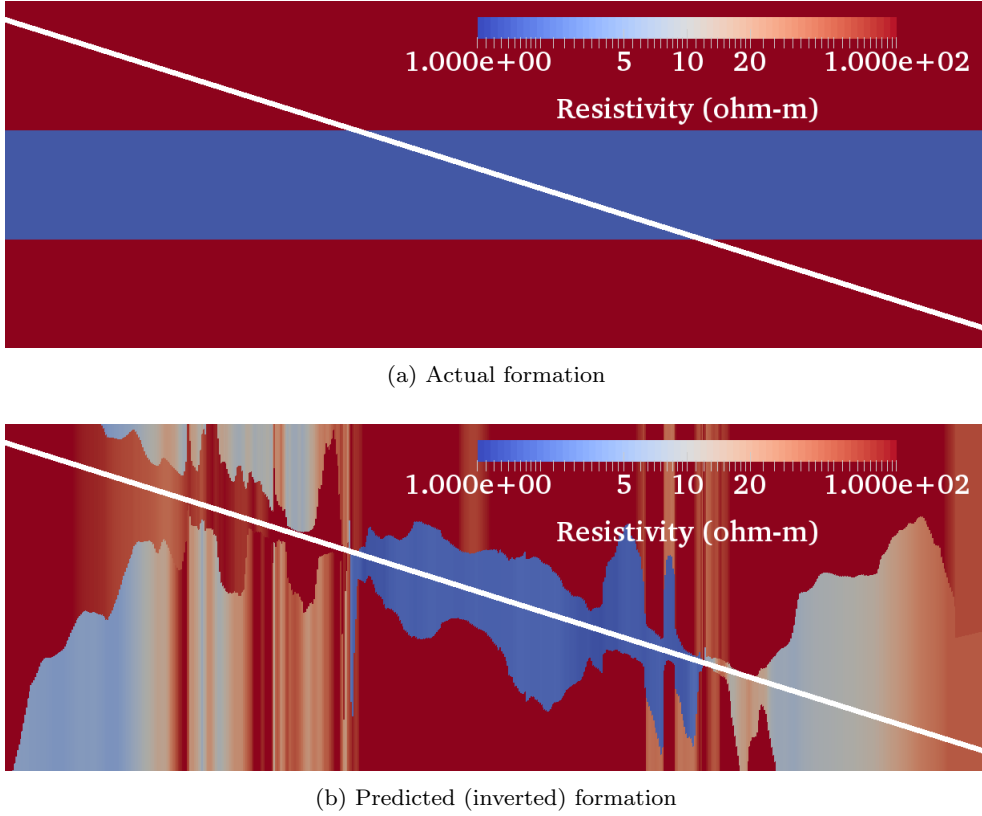


Fig. 11: Model problem 1. Comparison between actual and predicted (inverted) formation.

the result of the inner layer computations to the output $\mathcal{I}_\theta(\mathbf{x}_t)$, and $\mathbf{b}_a, \mathbf{b}_p$ are bias vectors allowing layers within the network to learn an offset. None of the weight matrices depend on the temporal component t and remain fixed, and the transition matrix \mathbf{W}_{aa} of the RNN is reset between processing two independent sequences.

The temporal nature of the process described in Equation (21) is better illustrated if operations are unfolded, as shown in Figure 28. Following this representation, an RNN can be interpreted not as cyclic, but as a standard network with one layer per time step and shared weights across time steps. It becomes clear that the network can be trained across many time steps using a variant of standard backpropagation algorithm, termed *backpropagation through time* [34,35].

From these first principles, many different flavors of RNNs have been successfully applied over time to temporal data. In this work, we make use of two significant advances in the field of RNNs, namely long-short term memory RNN (LSTM), and bidirectional recurrent neural network (BRNN).

LSTM networks [36] are similar to a standard RNN with one inner layer, but a so-called *memory cell* replaces each ordinary node in this layer. Each memory cell contains a node with a self-connected recurrent edge

of fixed weight one, ensuring that the gradient can be propagated across many time steps without vanishing or exploding. BRNNs contain two layers, both linked to input and output [37]. These two layers are different: the first has a recurrent connection from the past time steps while in the second the direction of recurrent of connections is reversed, performing computations backward along the sequence. More details about both architectures can be found in [38].

C Proposed Neural Network Architecture

The following is a listing of the neural network architecture built in this work in the Keras framework [39]:

```
i = Input(shape=input_shape)
x = LSTM(recurrent_output_size)(i)
x2 = Reshape((recurrent_output_size, 1))(x)
a = Conv1D(filters=nb_filter,
            kernel_size=3, activation='relu',
            kernel_initializer=glorot_normal(), padding='same')(x2)
d = Conv1D(filters=nb_filter,
            kernel_size=3, activation='relu',
            ,
```

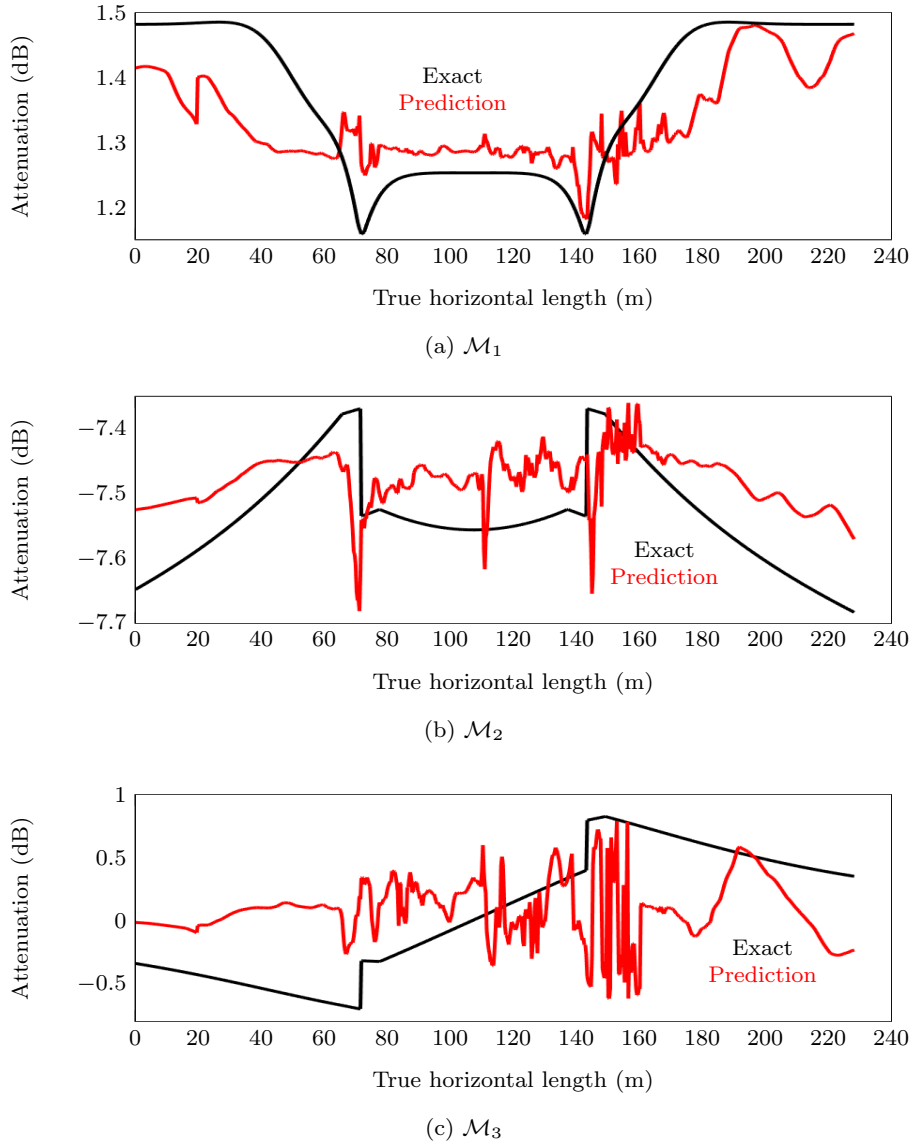


Fig. 12: Model problem 1. Comparison between exact and prediction of attenuation of the measurements.

```

kernel_initializer=glorot_normal(),
padding='same')(a)
x = Add()([x2,d])
x = MaxPooling1D(pool_length)(x)
a = Conv1D(filters=nb_filter,
kernel_size=3,activation='relu',
kernel_initializer=
glorot_normal(),padding='same')(
x)
d = Conv1D(filters=nb_filter,
kernel_size=3, activation='relu',
kernel_initializer=
glorot_normal(),padding='same')(
a)
x = Add()([x,d])
x = MaxPooling1D(pool_length)(x)
a = Conv1D(filters=nb_filter,
kernel_size=3, activation='relu',
kernel_initializer=
glorot_normal(),padding='same')(
x)
d = Conv1D(filters=nb_filter,
kernel_size=3, activation='relu',
kernel_initializer=
glorot_normal(),padding='same')(
x)
x = Add()([x,d])
x = MaxPooling1D(pool_length)(x)
x= Flatten(input_shape=input_shape)(
x)
y = Dense(num_outputs, activation='
sigmoid', kernel_initializer='
glorot_uniform')(x)

```

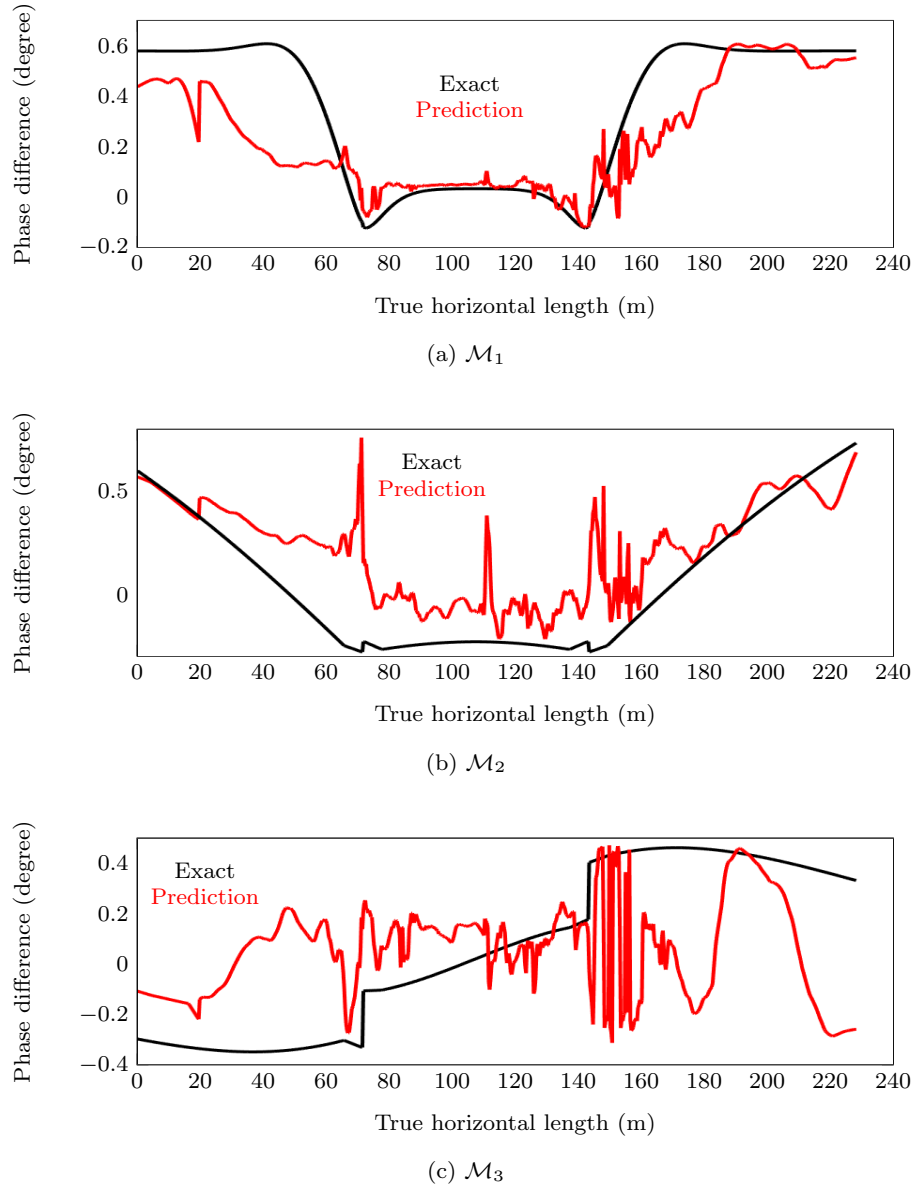


Fig. 13: Model problem 1. Comparison between exact and prediction of phase difference of the measurements.

References

1. Constable, S., Srnka, L.J.: An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. *Geophysics* **72**(2), WA3–WA12 (2007)
2. Bakr, S.A., Pardo, D., Mannseth, T.: Domain decomposition fourier fe method for the simulation of 3d marine csem measurements. *Journal of Computational Physics* **255**, 456–470 (2013)
3. Hardage, B.A.: Vertical seismic profiling. *The Leading Edge* **4**(11), 59–59 (1985)
4. Alvarez-Aramberri, J., Pardo, D.: Dimensionally adaptive hp-finite element simulation and inversion of 2D magnetotelluric measurements. *Journal of Computational Science* **18**, 95–105 (2017)
5. Davdycheva, S., Wang, T.: A fast modelling method to solve Maxwell’s equations in 1D layered biaxial anisotropic medium. *Geophysics* **76**(5), F293–F302 (2011)
6. Ijasana, O., Torres-Verdín, C., Preeg, W.E.: Inversion-based petrophysical interpretation of logging-while-drilling nuclear and resistivity measurements. *Geophysics* **78**(6), D473–D489 (2013)
7. Davdycheva, S., Homan, D., Minerbo, G.: Triaxial induction tool with electrode sleeve: FD modeling in 3D geometries. *Journal of Applied Geophysics* **67**, 98–108 (2004)
8. Shahriari, M., Rojas, S., Pardo, D., Rodríguez-Rozas, A., Bakr, S.A., Calo, V.M., Muga, I.: A numerical 1.5D method for the rapid simulation of geophysical resistivity measurements. submitted to *Journal of Computational Physics*, October 2017 (2017)
9. Pardo, D., Torres-Verdin, C.: Fast 1D inversion of logging-while-drilling resistivity measurements for the

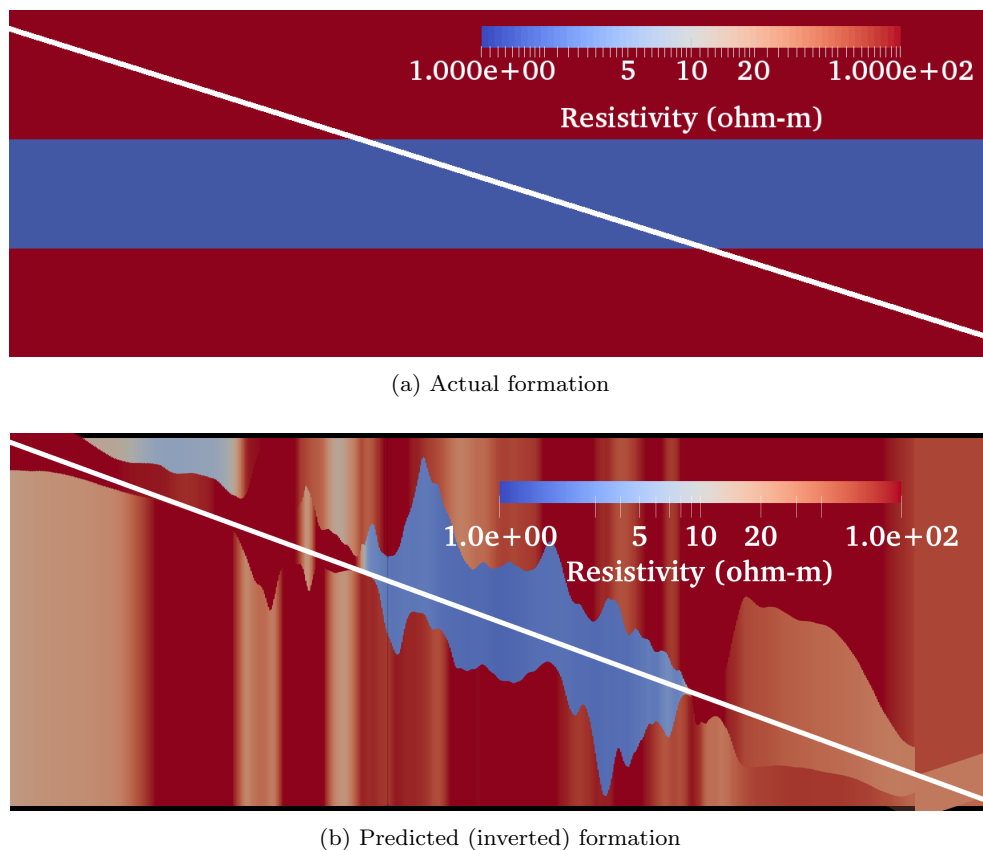


Fig. 14: Model problem 1. Comparison between actual and predicted (inverted) formation when using regularization.

- improved estimation of formation resistivity in high-angle and horizontal wells. *Geophysics* **80**(2), E111–E124 (2014)
10. Key, K.: 1D inversion of multicomponent, multifrequency marine CSEM data: Methodology and synthetic studies for resolving thin resistive layers. *Geophysics* **74**(2), F9–F20 (2009)
11. Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics (2005)
12. Vogel, C.: *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics (2002)
13. Watzenig, D.: Bayesian inference for inverse problems—statistical inversion. *Elektrotechnik & Informationstechnik* **124**, 240–247 (2007)
14. Ivakhnenko, A.G.: *Cybernetic Predicting Devices*. CCM Information Corporation (1973)
15. Dechter, R.: Learning while searching in constraint-satisfaction-problems. *Proceedings of the Fifth AAAI National Conference on Artificial Intelligence* pp. 178–183 (1986)
16. Aizenberg, I., Aizenberg, N.N., Vandewalle, J.P.L.: Multi-valued and universal binary neurons: theory, learning and applications. Springer Science & Business Media (2000)
17. Lu, L., Zheng, Y., Carneiro, G., Yang, L.: *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*. Springer, Switzerland (2017)
18. Yu, D., Deng, L.: *Automatic Speech Recognition: A Deep Learning approach*. Springer, London (2017)
19. Bhanu, B., Kumar, A.: *Deep Learning for Biometrics*. Springer, Switzerland (2017)
20. Bougher, B.B.: *Machine learning applications to geophysical data analysis*. Master's thesis, The University of British Columbia (2016)
21. Araya-Polo, M., Dahlke, T., Frogner, C., Zhang, C., Poggio, T., Hohl, D.: Automated fault detection without seismic processing. *Leading Edge* **36**(3), 208–214 (2017)
22. Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L.: Machine learning in geosciences and remote sensing. *Geoscience Frontiers* **7**(1), 3–10 (2016). Special Issue: Progress of Machine Learning in Geosciences
23. Hegde, C., Wallace, S., Gray, K.: Using trees, bagging, and random forests to predict rate of penetration during drilling. *Society of Petroleum Engineers* pp. 1–12 (2015)
24. Aulia, A., Rahman, A., Velasco, J.J.Q.: Strategic well test planning using random forest. *Society of Petroleum Engineers* pp. 1–23 (2014)
25. Bize-Forest, N., Lima, L., Baines, V., Boyd, A., Abbots, F., Barnett, A.: Using machine-learning for depositional facies prediction in a complex carbonate reservoir. *Society of Petrophysicists and Well-Log Analysts* pp. 1–11 (2018)
26. Wang, Y., S. W. Cheung, E.T.C., Efendiev, Y., Wang, M.: Deep multiscale model learning. *arXiv:1806.04830* (2018)

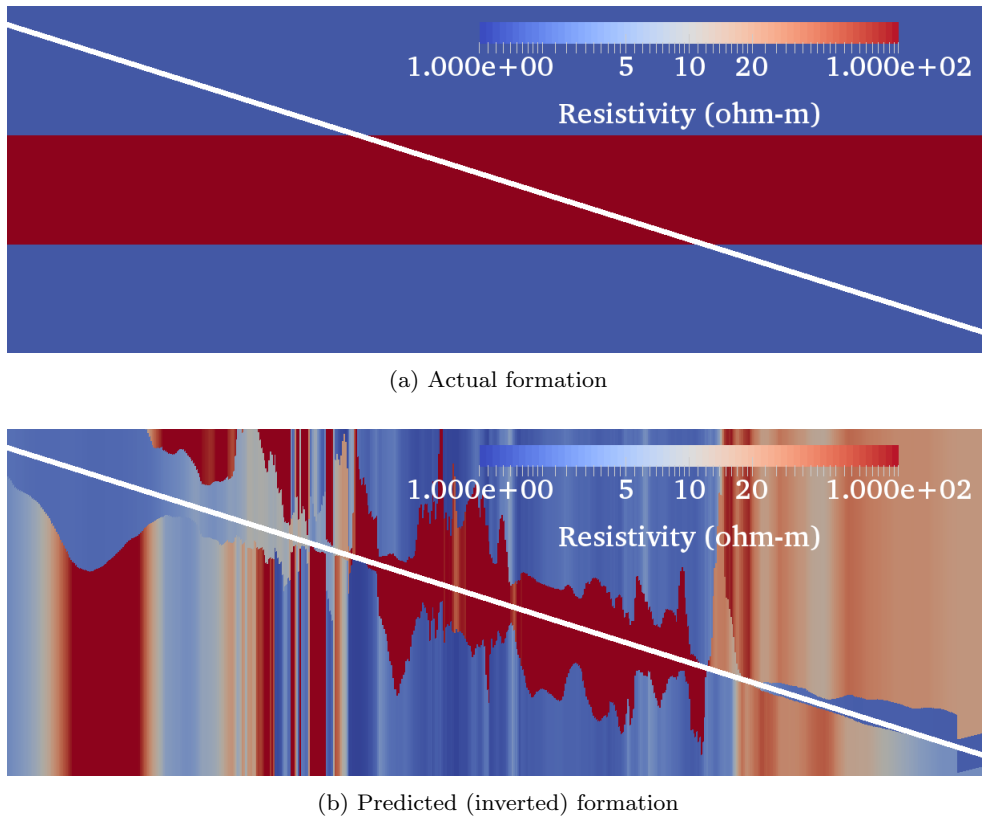


Fig. 15: Model problem 2. Comparison between actual and predicted (inverted) formation.

27. Higham, C.F., Higham, D.J.: Deep learning: An introduction for applied mathematicians. Computing Research Repository **abs/1801.05894** (2018)
28. Key, K.: 1D inversion of multicomponent, multifrequency marine CSEM data: Methodology and synthetic studies for resolving thin resistive layers. *Geophysics* **74**(2), F9–F20 (2009)
29. Hornik, K.: Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**(2), 251–257 (1991)
30. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1. pp. 318–362. MIT Press, Cambridge, MA, USA (1986)
31. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998)
32. Hopfield, J.J.: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America* **79**(8), 2554–2558 (1982)
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. *arXiv:1512.03385* (2015)
34. Werbos, P.J.: Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* **78**(10), 1550–1560 (1990)
35. Hochreiter, S., Bengio, Y., Frasconi, P.: Gradient flow in recurrent nets: The difficulty of learning long-term dependencies. In: *Field Guide to Dynamical Recurrent Networks*. IEEE Press (2001)
36. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997)
37. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* **45**(11), 2673–2681 (1997)
38. Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning (2015)
39. Chollet, F.: Keras. <https://github.com/fchollet/keras> (2015)

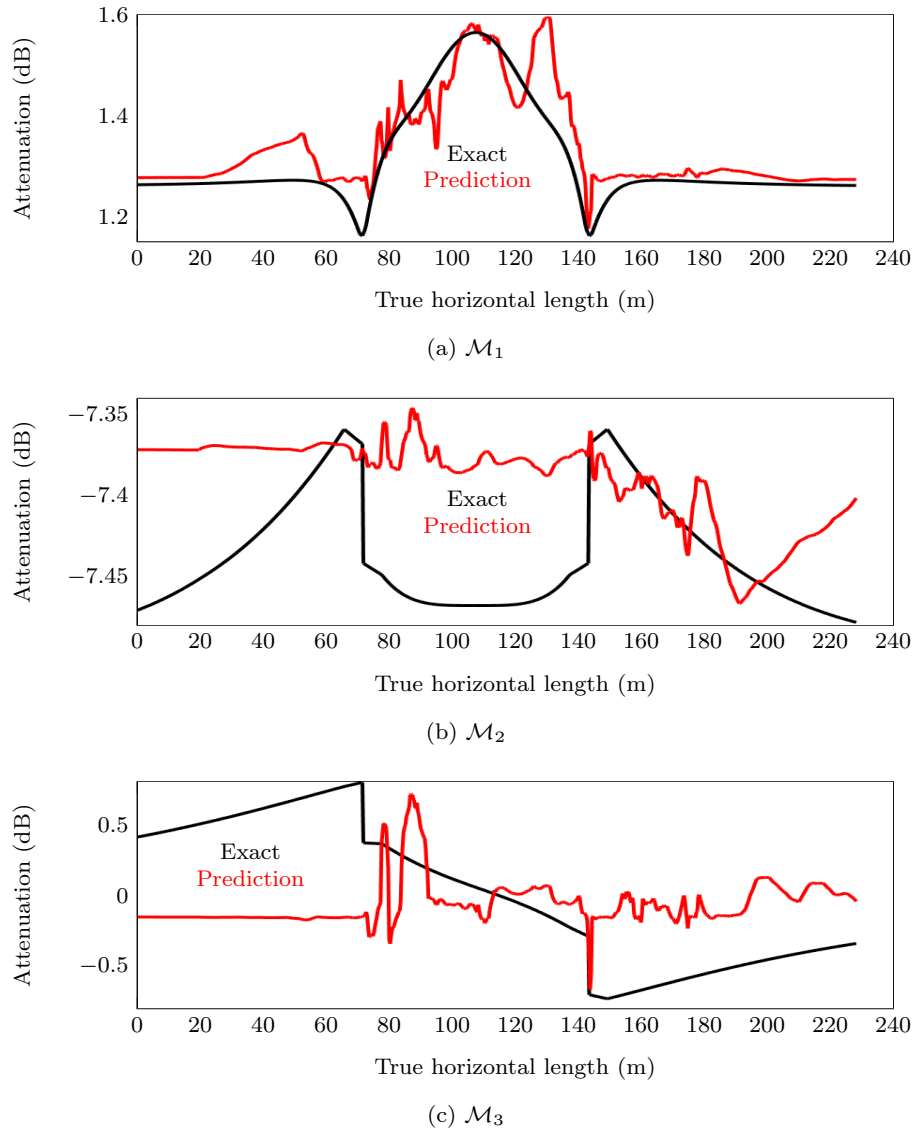


Fig. 16: Model problem 2. Comparison between exact and prediction of attenuation of the measurements.

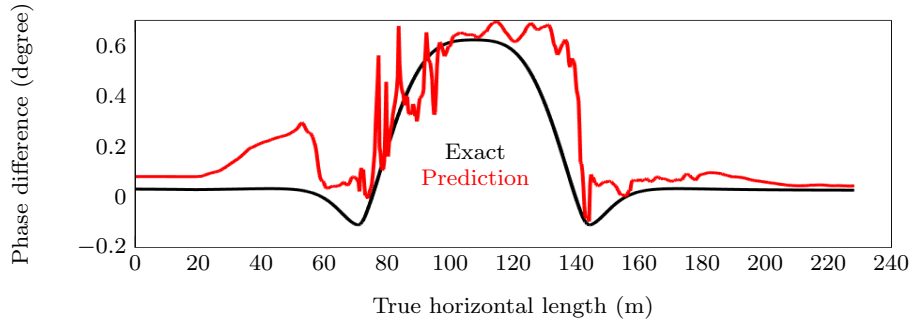
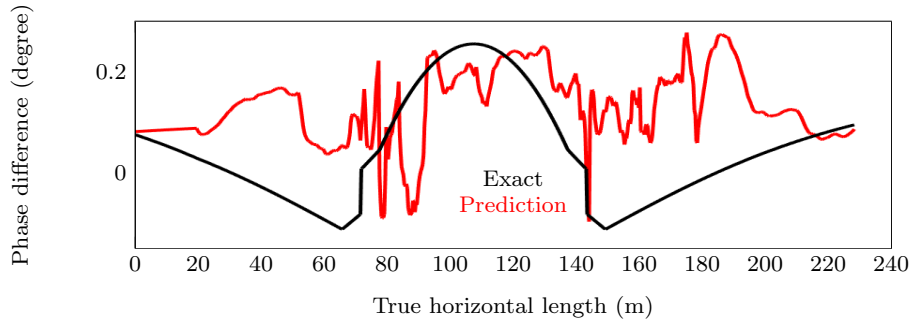
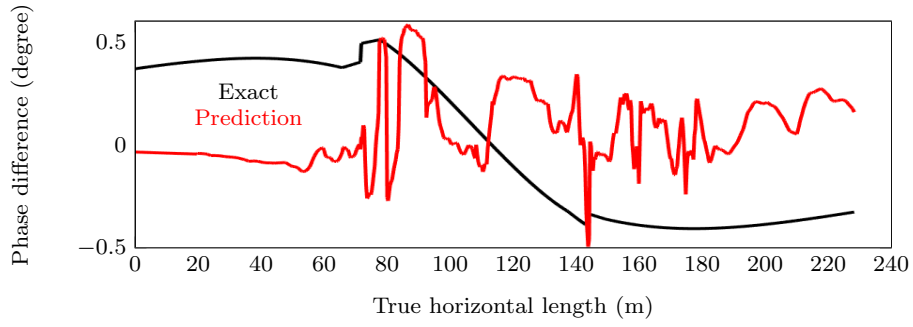
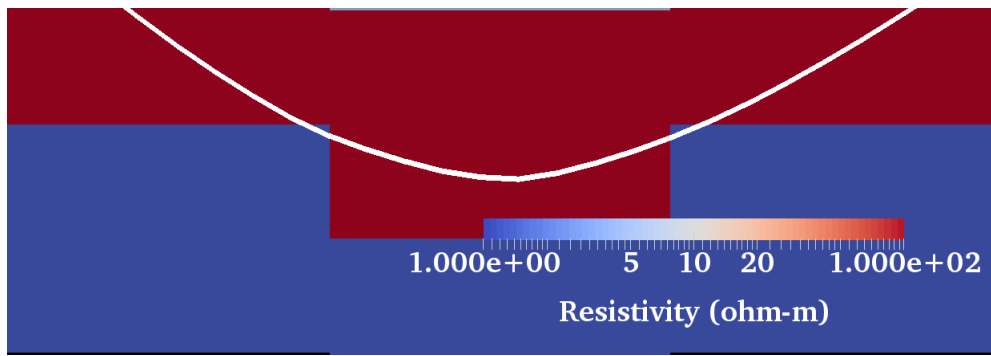
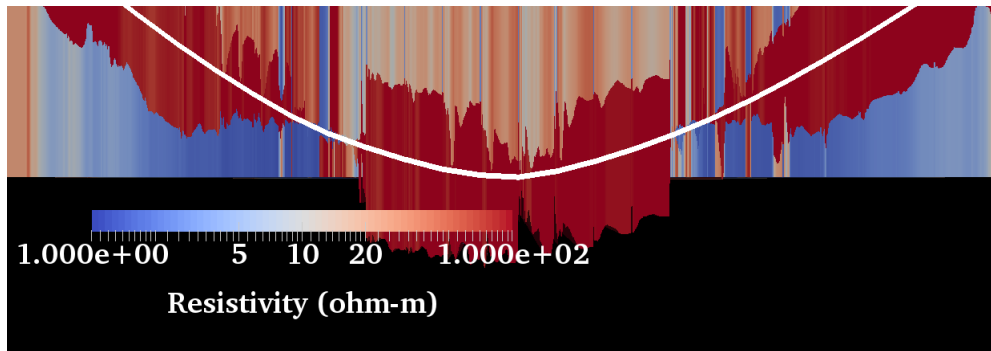
(a) \mathcal{M}_1 (b) \mathcal{M}_2 (c) \mathcal{M}_3

Fig. 17: Model problem 2. Comparison between exact and prediction of phase difference of the measurements.



(a) Actual formation



(b) Predicted (inverted) formation

Fig. 18: Model problem 3. Comparison between actual and predicted (inverted) formation.

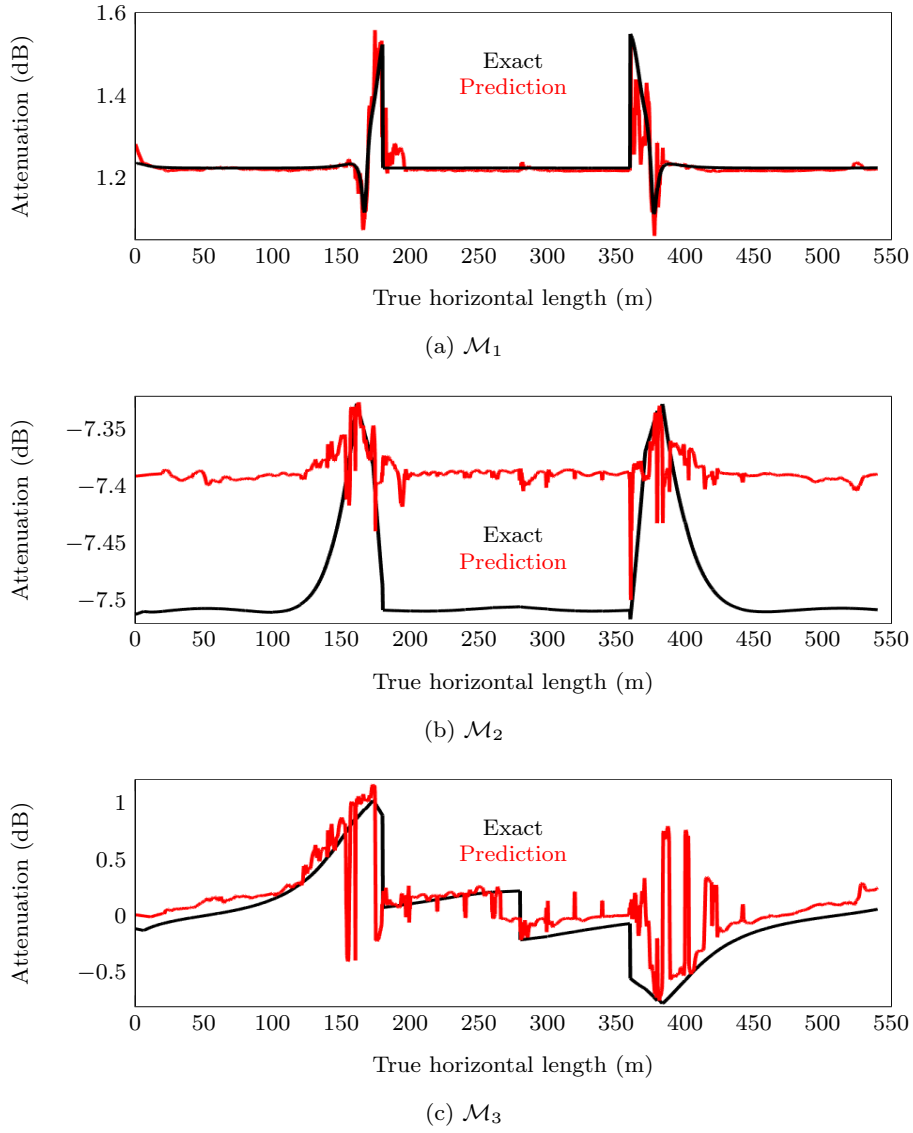


Fig. 19: Model problem 3. Comparison between exact and prediction of attenuation of the measurements.

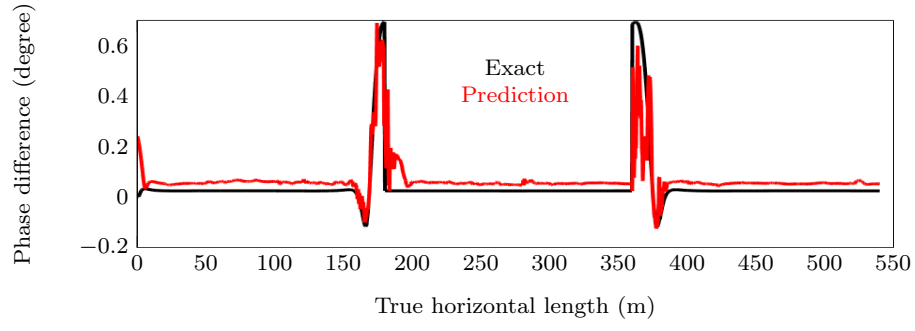
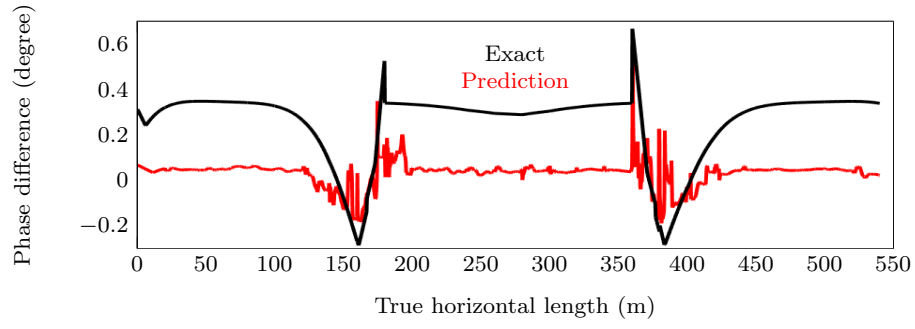
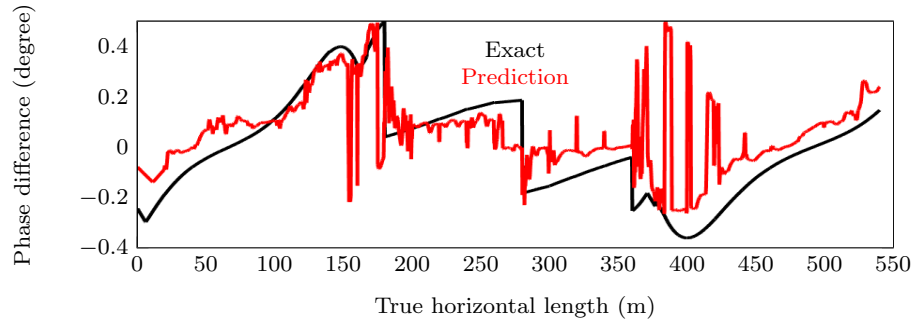
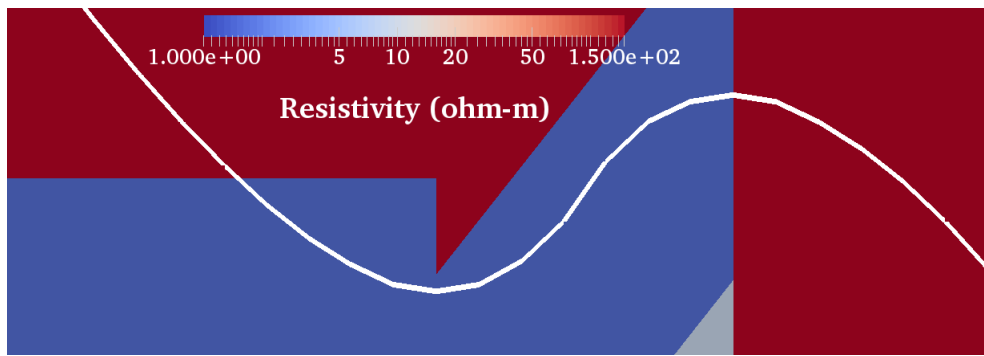
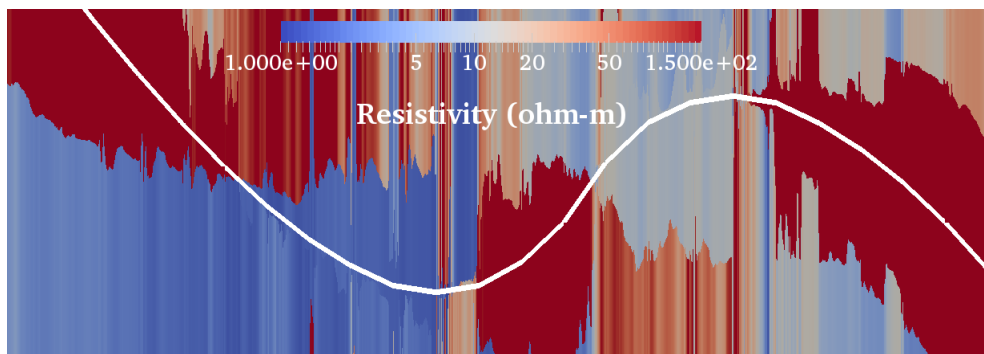
(a) \mathcal{M}_1 (b) \mathcal{M}_2 (c) \mathcal{M}_3

Fig. 20: Model problem 3. Comparison between exact and prediction of phase difference of the measurements.



(a) Actual formation



(b) Predicted (inverted) formation

Fig. 21: Model problem 4. Comparison between actual and predicted (inverted) formation.

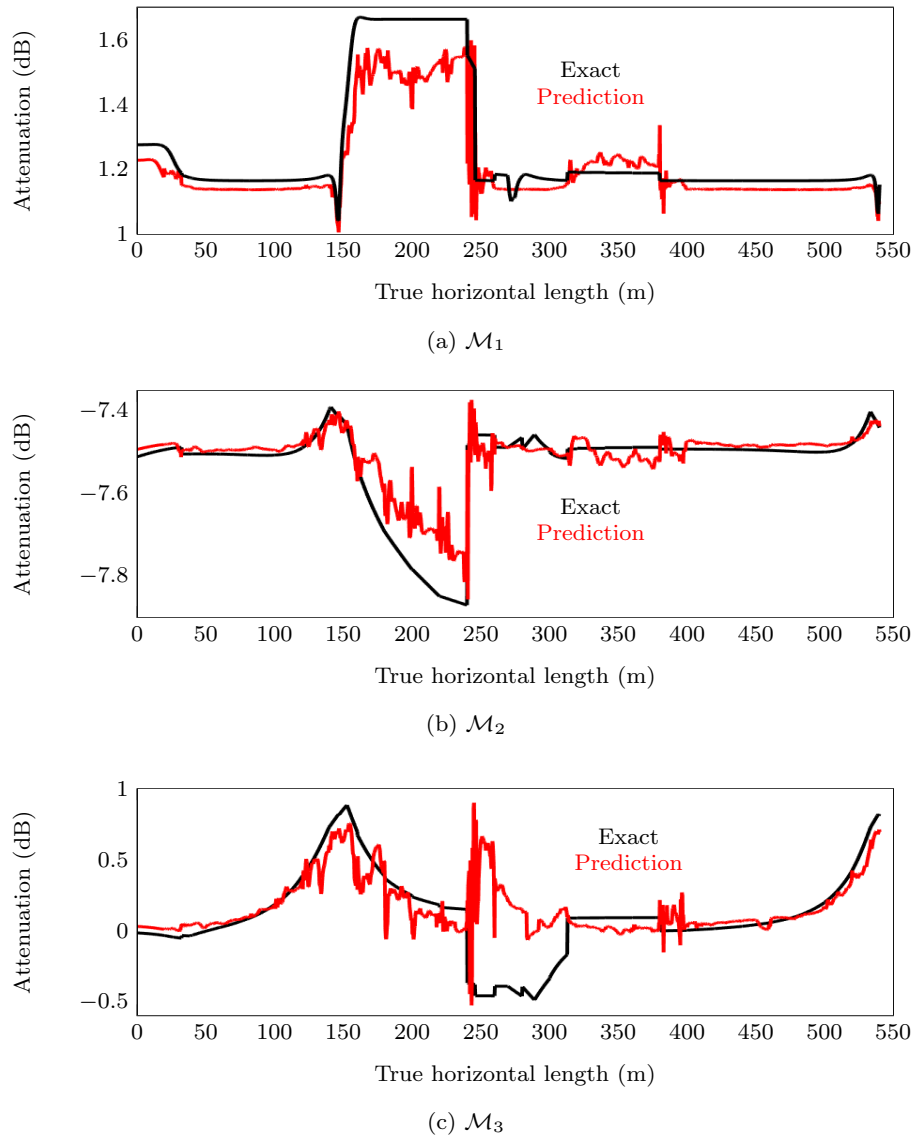


Fig. 22: Model problem 4. Comparison between exact and prediction of attenuation of the measurements.

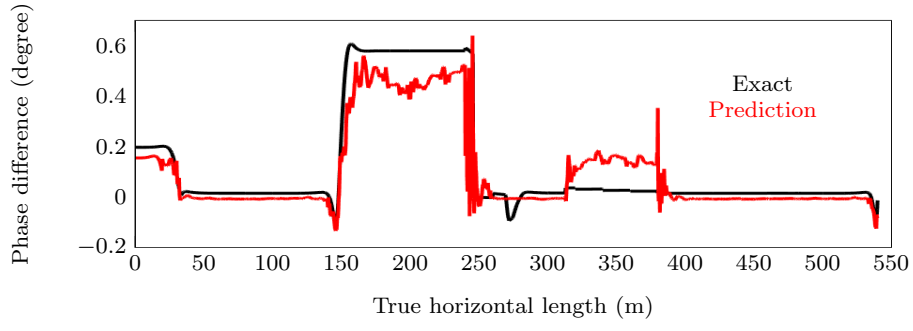
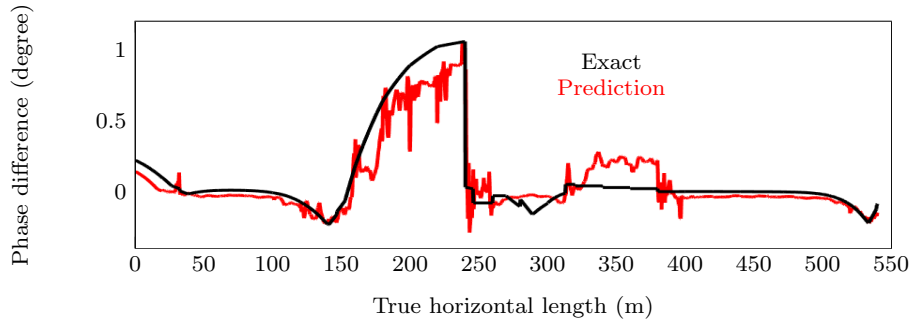
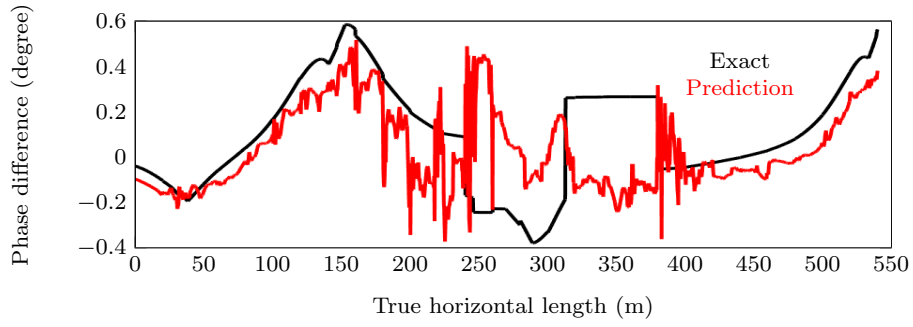
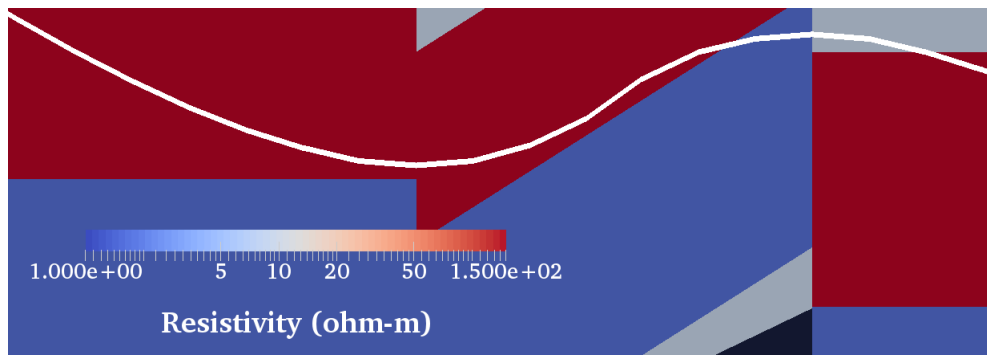
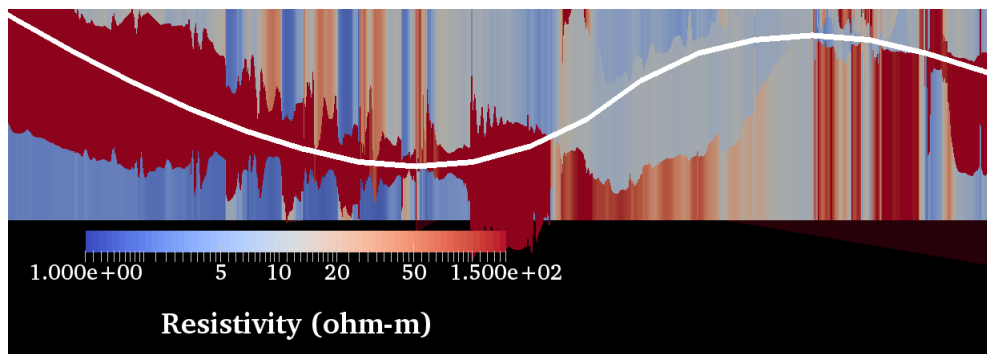
(a) \mathcal{M}_1 (b) \mathcal{M}_2 (c) \mathcal{M}_3

Fig. 23: Model problem 4. Comparison between exact and prediction of phase difference of the measurements.



(a) Actual formation



(b) Predicted (inverted) formation

Fig. 24: Model problem 4. Comparison between actual and predicted (inverted) formation.

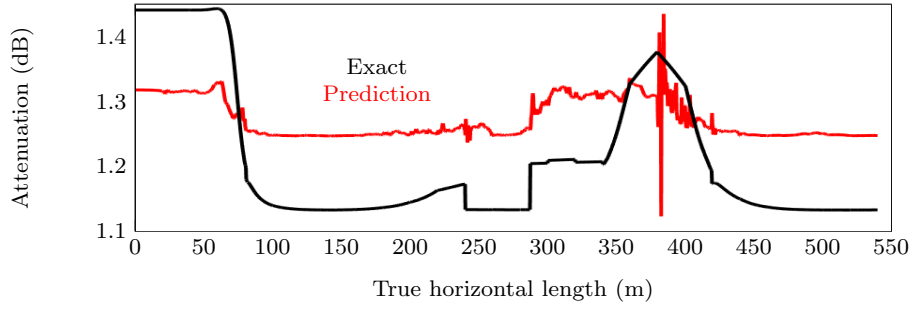
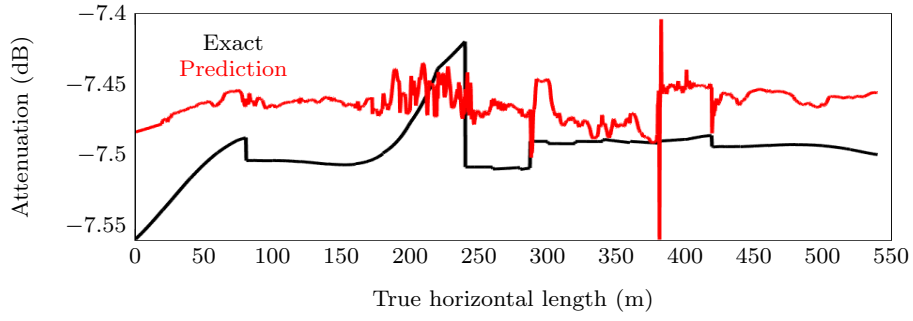
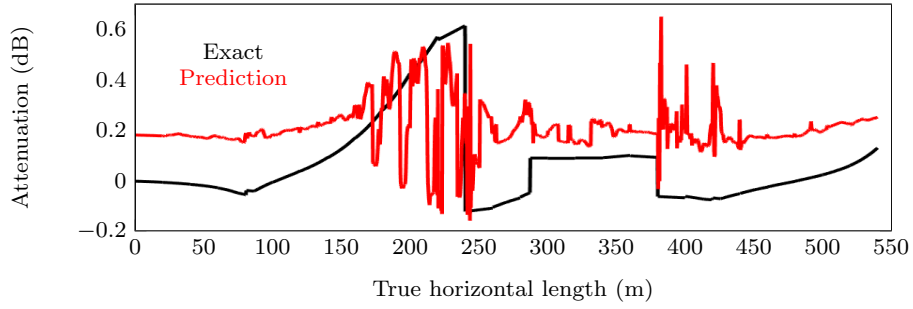
(a) \mathcal{M}_1 (b) \mathcal{M}_2 (c) \mathcal{M}_3

Fig. 25: Model problem 5. Comparison between exact and prediction of attenuation of the measurements.

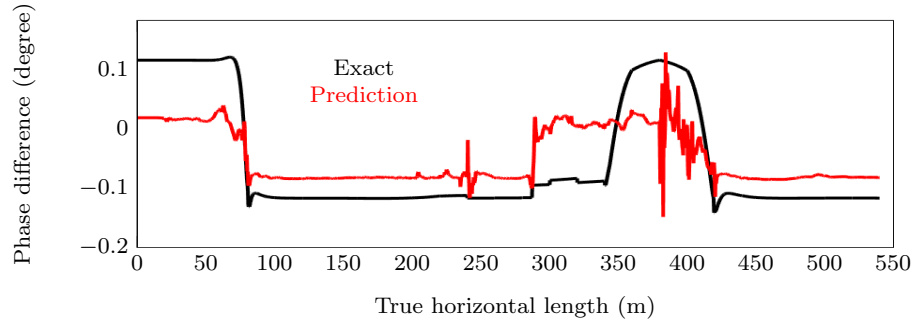
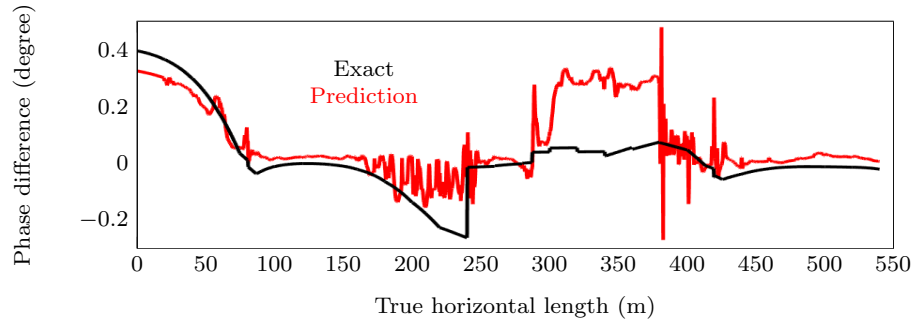
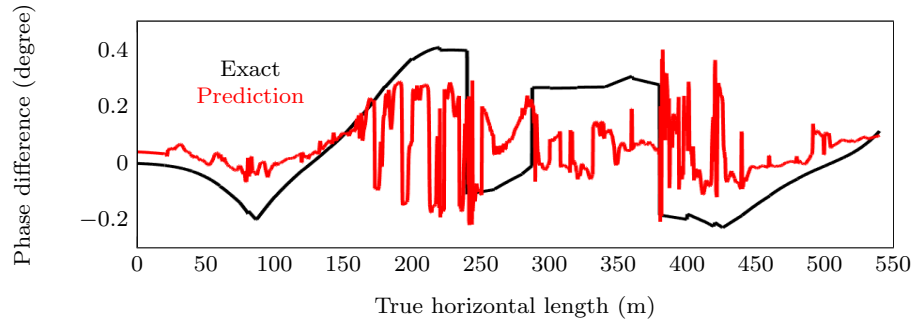
(a) \mathcal{M}_1 (b) \mathcal{M}_2 (c) \mathcal{M}_3

Fig. 26: Model problem 5. Comparison between exact and prediction of phase difference of the measurements.

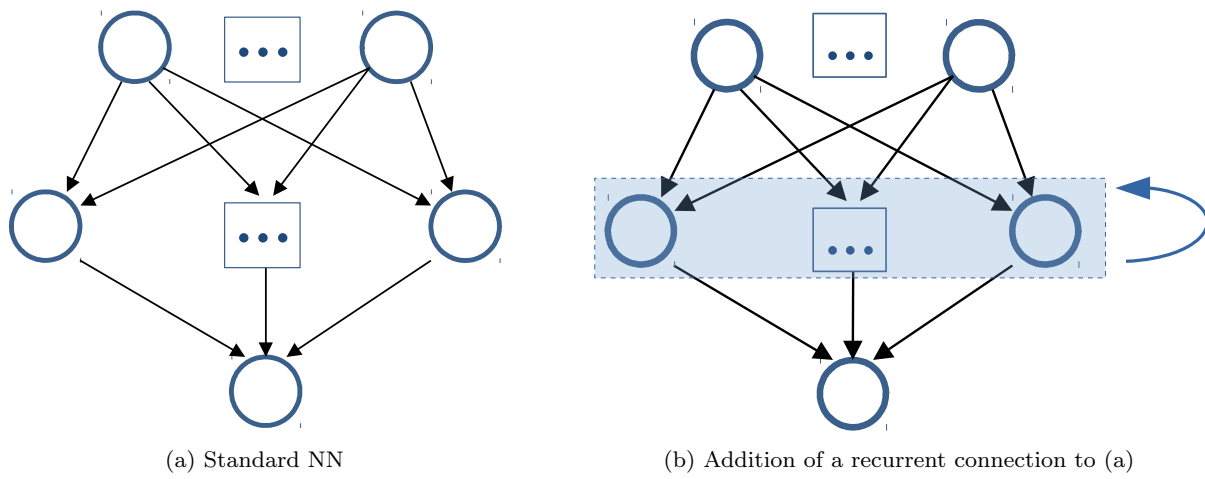


Fig. 27: Comparison between NN and RNN

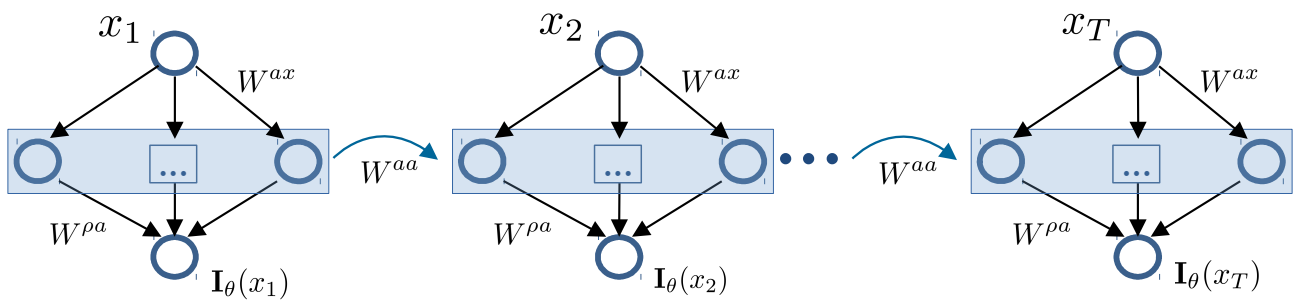


Fig. 28: RNN with computations unfolded through time.